

N° d'ordre

THÈSES

présentées

A LA FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE PARIS

pour obtenir

LE GRADE DE DOCTEUR ES SCIENCES MATHÉMATIQUES

par

Claude GILORMINI

Première thèse

APPROXIMATION RATIONNELLE AVEC CONTRAINTES

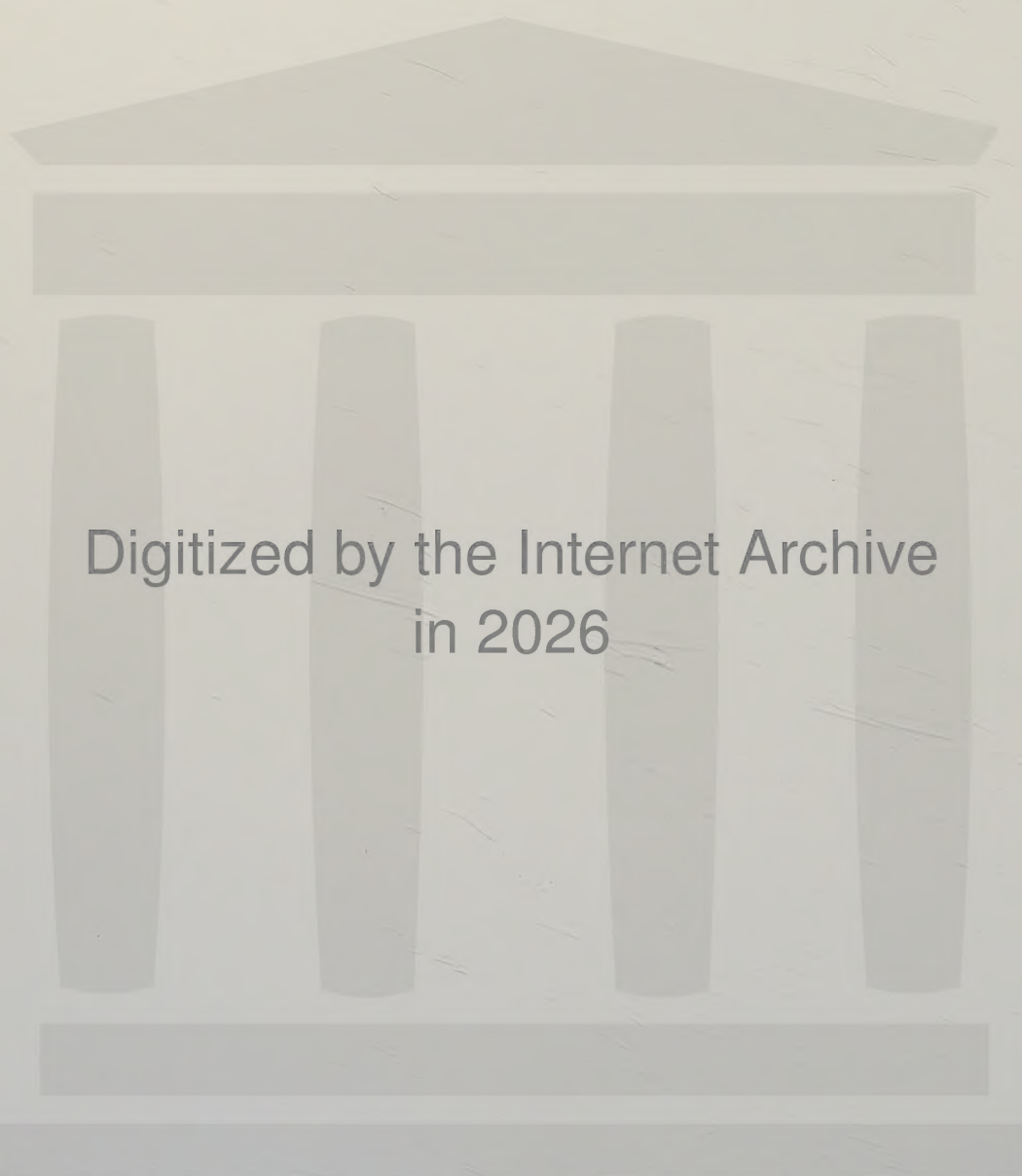
Deuxième thèse

THEOREMES ASYMPTOTIQUES DANS LES CHAÎNES DE MARKOV

Soutenues le 23 AVRIL 1968, devant la Commission d'examen

M. P. de POSSEL Président

M. J. L. LIONS
M. J. NEVEU Examinateurs



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41554025_0041

THÈSES

présentées

A LA FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE PARIS

pour obtenir

LE GRADE DE DOCTEUR ES SCIENCES MATHÉMATIQUES

par

Claude GILORMINI

Première thèse

APPROXIMATION RATIONNELLE AVEC CONTRAINTES

Deuxième thèse

THEOREMES ASYMPTOTIQUES DANS LES CHAÎNES DE MARKOV

Soutenues le

, devant la Commission d'examen

M. R. de POSSEL *Président*

M. J. L. LIONS

Examineurs

M. J. NEVEU

THÈSES

présentées

à la Faculté des Sciences de l'Université de Paris

par

le Docteur en Sciences Mathématiques

et

le Docteur en Sciences Physiques

et

le Docteur en Sciences Médicales



et

le Docteur en Sciences Littéraires

et

le Docteur en

Sciences

et

le Docteur en

Sciences

TABLE DES MATIERES

INTRODUCTION	1
CHAPITRE 0 : Résultats préliminaires	
1- Théorème de Carathéodory	5
2- Compatibilité d'un système d'inéquations linéaires	7
3- Propriétés des sous-espaces de Haar	8
CHAPITRE I : Approximation rationnelle avec contraintes générales	
1- Introduction	15
2- Existence d'un élément optimal	19
3- Convexité de l'ensemble des éléments optimaux	25
4- Extension du théorème de La Vallée-Poussin	26
5- Caractérisation des éléments optimaux	28
6- Propriétés des coefficients de l'élément optimal	31
7- Unicité	34
8- Propriétés d'alternance	35
9- Cas d'une norme faisant intervenir les dérivées	36
CHAPITRE II : Approximation rationnelle avec contraintes du type égalités et inégalités linéaires	
1- Introduction	43
2- Existence d'un élément optimal	45
3- Caractérisation des éléments optimaux	46
4- Unicité	59
5- Propriétés d'alternance	69
CHAPITRE III : Approximation rationnelle avec contraintes du type égalité	
1- Introduction	71
2- Existence d'un élément optimal	73
3- Caractérisation des éléments optimaux	74
4- Propriétés d'alternance	75
5- Unicité	79
6- Unicité forte	84
7- Continuité	89
8- Retour sur le cas des contraintes inégalités	92

CHAPITRE IV : Approximation rationnelle avec conditions de noeuds

1- Introduction	95
2- Existence d'un élément optimal	97
3- Caractérisation des éléments optimaux	98
4- Propriétés d'alternance	100
5- Unicité	105
6- Unicité forte	112
7- Continuité	114
8- Cas des noeuds extérieurs à l'intervalle I	115
9- Retour sur le cas des contraintes inégalités	117

CHAPITRE V : Etude de quelques cas particuliers

1- Fractions rationnelles ordinaires avec conditions de noeuds	121
2- Fractions rationnelles trigonométriques avec conditions de noeuds	127
3- Fractions rationnelles ordinaires avec conditions d'inégalité	134
4- Fractions rationnelles trigonométriques avec conditions d'inégalité	138

BIBLIOGRAPHIE

145

INTRODUCTION

Le présent travail a pour but l'étude de l'approximation au sens de Tchebysheff d'une fonction continue sur un compact par des fractions rationnelles généralisées dont les coefficients sont soumis à certaines contraintes.

On considère l'espace $C[X]$ des fonctions réelles continues sur un espace compact X et on le munit de la norme :

$$\|g\| = \max_{x \in X} |g(x)|$$

On désigne par $W_{m,n}$ la classe des "fractions rationnelles généralisées" :

$$R(x) = \sum_0^m a_j g_j(x) / \sum_0^n b_j h_j(x)$$

telles que :

$$(a_0, a_1, \dots, a_m, b_0, b_1, \dots, b_n) \in \Omega$$

$$\sum b_j h_j(x) > 0 \quad \text{sur } X$$

où Ω est un certain ensemble donné dans R^{m+n+2} et où les fonctions $h_j(x), g_j(x)$ sont données dans $C[X]$.

Etant donné une fonction $f \in C[X]$, on cherche dans $W_{m,n}$ un élément R^* (qui sera dit optimal) tel que :

$$\|f - R^*\| = \inf_{W_{m,n}} \|f - R\|$$

Dans le chapitre I, on étudie le problème sous la forme générale ci-dessus. On donne en particulier une condition suffisante d'existence d'un élément optimal qui généralise le résultat démontré par BOEHM dans le cas sans contraintes ($\Omega = R^{m+n+2}$) et, dans le cas où Ω est convexe, une condition nécessaire et suffisante pour qu'un élément R de $W_{m,n}$ soit optimal.

Dans le chapitre II, on étudie le cas où les contraintes sont des égalités et inégalités linéaires, c'est-à-dire le cas où Ω est défini par les conditions :

$$\sum_0^m \alpha_{ij} a_j + \sum_0^n \beta_{ij} b_j = c_i \quad (i = 1, \dots, K)$$

$$\sum_0^m \alpha_j \alpha_j(t) + \sum_0^n b_j \beta_j(t) \geq c(t) \quad t \in G$$

En plus des résultats déduits du chapitre précédent, on obtient en particulier une nouvelle forme du théorème de caractérisation de l'élément optimal et une condition nécessaire et suffisante pour qu'il y ait unicité pour tout couple de fonctions $f(x)$, $c(t)$. Ces résultats généralisent ceux donnés par GOLSHTEIN dans le cas polynomial.

Dans le chapitre III, on considère le cas où les contraintes sont seulement des égalités linéaires :

$$\sum_0^m \alpha_{ij} a_j + \sum_0^n \beta_{ij} b_j = c_i \quad (i = 1, \dots, K)$$

On énonce tout d'abord diverses formes du théorème de caractérisation de l'élément optimal (analogues à celles données par CHENEY dans le cas sans contraintes). On étudie ensuite les propriétés d'alternance, portant sur le nombre de points (appelés (e)-points) où $f-R$ atteint son module maximum avec des signes alternés. On en déduit une nouvelle forme du théorème d'unicité (analogue au résultat donné par LOEB dans le cas sans contraintes). On généralise enfin les théorèmes d'unicité forte et de continuité (démontrés par CHENEY dans le cas sans contraintes). On montre que sous certaines conditions, on a :

$$\|f-R\| \geq \|f-R^*\| + \gamma \|R-R^*\| \quad \forall R \in W_{m,n}$$

et :

$$\|R^*-R_0^*\| < \beta \|f-f_0\|$$

où R^* et R_0^* sont respectivement optimaux pour f et f_0 .

Dans le chapitre IV, on considère le cas où les contraintes sont des conditions de noeuds, c'est-à-dire où l'on impose aux fractions $R(x)$ de prendre des valeurs données en des points donnés (noeuds).

On démontre des propriétés d'alternance analogues à celles démontrées par PASZKOWSKI dans le cas polynomial (ces propriétés font intervenir le nombre de (e)-points et le nombre de noeuds situés entre deux (e)-points consécutifs). On en déduit ensuite, comme au chapitre précédent, des théorèmes sur l'unicité, l'unicité forte et la continuité.

Enfin, dans le cas où tous les noeuds sont extérieurs à X , on vérifie que les résultats du présent chapitre et ceux du chapitre précédent deviennent analogues.

Dans le chapitre V, enfin, on considère quelques cas particuliers des problèmes étudiés dans les chapitre II et IV.

On considère d'abord des fractions rationnelles "ordinaires" :

$$R(x) = \frac{a_0 + a_1 x + \dots + a_m x^m}{b_0 + b_1 x + \dots + b_n x^n}$$

soumises, soit à des conditions de noeuds, soit à des inégalités du type :

$$V(x) \leq R(x) \leq W(x) \quad \forall x \in [a, b]$$

(où $V(x)$ et $W(x)$ sont deux fonctions données).

On considère ensuite les mêmes problèmes avec des fractions rationnelles "trigonométriques" :

$$R(\theta) = \frac{a_0 + \sum_1^m (a_j \cos j\theta + b_j \sin j\theta)}{c_0 + \sum_1^n (c_j \cos j\theta + d_j \sin j\theta)}$$

Je suis heureux de pouvoir exprimer ici mes remerciements et ma profonde gratitude à Monsieur de POSSEL qui m'a suivi et encouragé au cours de mes recherches.

Ma reconnaissance va également à Monsieur J.L. LIONS dont les conseils m'ont permis d'améliorer et de généraliser mes résultats.

Je remercie Monsieur J. NEVEU qui m'a fait l'honneur de participer à mon jury.

Enfin je remercie tous ceux qui à l'imprimerie de l'Institut Blaise Pascal ont apporté le plus grand soin à la réalisation pratique de cette thèse.

RESULTATS PRELIMINAIRES

Ce chapitre contient un certain nombre de résultats auxiliaires dont nous aurons besoin par la suite et dont nous rappelons les démonstrations.

On trouvera dans le paragraphe 1 un théorème sur l'enveloppe convexe d'un compact de R^N .

Le paragraphe 2 contient une condition nécessaire et suffisante pour qu'un système d'inéquations linéaires soit incompatible.

Le paragraphe 3 donne plusieurs propriétés des sous-espaces de Haar de $C[I]$, espace des fonctions continues sur un intervalle I .

1 - Théorème de Carathéodory.

Soit A un ensemble contenu dans R^N . On appelle *enveloppe convexe* de A et on note $H(A)$ l'ensemble des combinaisons linéaires convexes finies d'éléments de A c'est-à-dire les combinaisons linéaires finies de la forme :

$$g = \sum \theta_i f_i$$

avec :

$$f_i \in A, \theta_i \geq 0, \sum \theta_i = 1$$

Théorème 1 : Tout point de l'enveloppe convexe d'un ensemble A contenu dans R^N s'exprime comme combinaison linéaire convexe d'au plus $(N+1)$ éléments de A c'est-à-dire :

$$g = \sum_{i=0}^N \theta_i f_i \quad \text{avec} \quad f_i \in A, \theta_i \geq 0, \sum \theta_i = 1$$

Soit $g \in H(A)$. D'après la définition de l'enveloppe convexe, on a :

$$(1) \quad g = \sum_{i=0}^k \theta_i f_i \quad \text{avec} \quad \theta_i \geq 0, \quad \sum \theta_i = 1$$

Supposons k minimum, c'est-à-dire : $\theta_i > 0 \quad (i=0, 1, \dots, k)$

La relation (1) peut encore s'écrire :

$$(2) \quad \sum_{i=0}^k \theta_i (g - f_i) = 0$$

ce qui prouve que les vecteurs $g_i = g - f_i \quad (i=0, 1, \dots, k)$ sont dépendants.

Supposons $k > N$. Alors les vecteurs $g_1, g_2, \dots, g_k$ sont dépendants (puisqu'ils appartiennent à R^N). On a donc :

$$(3) \quad \sum_{i=1}^k \alpha_i g_i = 0$$

où les α_i sont des nombres réels non tous nuls.

De (2) et (3), on déduit :

$$(4) \quad \sum_{i=0}^k (\theta_i + \lambda \alpha_i) g_i = 0 \quad \text{où l'on pose : } \alpha_0 = 0$$

On peut choisir λ de manière à ce que les coefficients $\theta_i + \lambda \alpha_i$ soient tous positifs ou nuls, l'un d'entre eux étant nul.

(4) devient alors :

$$\sum_{i=0}^k (\theta_i + \lambda \alpha_i) g = \sum_{i=0}^k (\theta_i + \lambda \alpha_i) f_i$$

ou, en tenant compte de l'inégalité : $\sum (\theta_i + \lambda \alpha_i) \geq \theta_0 > 0$

$$(5) \quad g = \frac{\sum_{i=0}^k (\theta_i + \lambda \alpha_i) f_i}{\sum_{i=0}^k (\theta_i + \lambda \alpha_i)}$$

Comme l'un des coefficients $(\theta_i + \lambda \alpha_i)$ est nul, la relation (5) montre que g est une combinaison linéaire convexe de k éléments seulement, ce qui contredit l'hypothèse k minimum. On ne peut donc pas avoir $k > N$, ce qui démontre le théorème.

2- Compatibilité d'un système d'inéquations linéaires.

Dans R^N , nous noterons $\langle x, y \rangle$ le produit scalaire :

$$\langle x, y \rangle = \sum_{i=1}^N x_i y_i$$

Théorème 2 : soit U un ensemble compact dans R^N . Pour que l'origine de R^N appartienne à l'enveloppe convexe de U , il faut et il suffit que le système d'inéquations $\langle u, x \rangle > 0 \quad \forall u \in U$ soit incompatible. (CHENEY [1966])

ou : il faut et il suffit que $\min_{u \in U} \langle u, x \rangle \leq 0$ pour tout $x \in R^N$

1°) Supposons que $0 \in H(U)$. On a alors :

$$0 = \sum_{i=1}^m \lambda_i u_i \quad \text{avec} \quad \lambda_i \geq 0, \quad \sum \lambda_i = 1, \quad u_i \in U$$

D'où l'on déduit

$$\sum_{i=1}^m \lambda_i \langle u_i, z \rangle = 0 \quad \text{quel que soit} \quad z \in R^N$$

Cette relation prouve que le système d'inéquations suivant est incompatible :

$$\langle u_i, z \rangle > 0 \quad (i=1, \dots, m)$$

Il en résulte bien entendu que le système d'inéquations :

$$\langle u, z \rangle > 0 \quad \forall u \in U \quad \text{est incompatible.}$$

2°) Supposons que $0 \notin H(U)$.

U étant compact, il en est de même pour $H(U)$. Puisque l'origine n'appartient pas à $H(U)$, il existe un élément $z \in H(U)$ dont la norme est minimum :

$$\|z\| = \inf_{u \in H(U)} \|u\|$$

Soit v un élément arbitraire de U . Quel que soit $\theta \in [0, 1]$, $\theta v + (1 - \theta)z$ est élément de $H(U)$ et donc :

$$\|\theta v + (1 - \theta)z\|^2 - \|z\|^2 \geq 0$$

ou encore :

$$(6) \quad \theta^2 \|v - z\|^2 + 2\theta \langle v - z, z \rangle \geq 0$$

Pour θ assez petit, la relation (6) ne peut être vérifiée que si l'on a :

$$\langle v - z, z \rangle \geq 0$$

c'est-à-dire :

$$(7) \quad \langle v, z \rangle \geq \langle z, z \rangle > 0$$

Le point v étant arbitraire dans U , la relation (7) montre que le système d'inéquations $\langle u, x \rangle > 0 \quad \forall u \in U$ admet z pour solution.

Ce qui termine la démonstration du théorème.

3- Propriétés des sous-espaces de Haar.

Soit $I = [a, b]$ un intervalle compact de R et soit $C[I]$ l'espace vectoriel des fonctions à valeurs réelles continues sur I .

On dit qu'un sous-espace vectoriel M de dimension finie de $C[I]$ est un sous-espace de Haar si tout élément $\Phi \in M$ non identiquement nul a au plus $\dim(M) - 1$ zéros dans I .

Rappelons que si $\Phi_1(x), \Phi_2(x), \dots, \Phi_N(x)$ est une base de M , cette condition est équivalente à la relation suivante :

$$(8) \quad \begin{vmatrix} \Phi_1(x_1) & \dots & \Phi_1(x_N) \\ \dots & \dots & \dots \\ \Phi_N(x_1) & \dots & \Phi_N(x_N) \end{vmatrix} \neq 0 \text{ quels que soient } x_1, \dots, x_N \\ \text{distincts dans } I$$

Théorème 3 : soit M un sous-espace de Haar de $C[I]$ de dimension finie N .

Tout élément $\Phi \neq 0$ de M a au plus $(N-1)$ zéros dans I , en comptant deux fois les zéros intérieurs à I où Φ ne change pas de signe.

RICE [1964, lemme 3-2], KARLIN - STUDDEN [1966].

Soit Φ un élément de M . Notons $x_1, \dots, x_p$ les zéros simples de Φ (où Φ change de signe) et $y_1, \dots, y_q$ les zéros doubles (où Φ ne change pas de signe). Désignons par γ_j le signe de Φ au voisinage du point y_j .

Considérons alors le système d'équations linéaires suivant :

$$(9) \quad \begin{cases} a_1 \Phi_1(x_j) + \dots + a_N \Phi_N(x_j) = 0 & j = 1, \dots, p \\ a_1 \Phi_1(y_j) + \dots + a_N \Phi_N(y_j) = -\gamma_j \varepsilon & j = 1, \dots, q \end{cases}$$

D'après (8), ce système est exactement de rang $p+q \leq N-1$.

Donc, quel que soit ε , il existe dans M un élément Ψ_ε non identiquement nul tel que :

$$\begin{cases} \Psi_\varepsilon(x_j) = 0 & j = 1, \dots, p \\ \Psi_\varepsilon(y_j) = -\gamma_j \varepsilon & j = 1, \dots, q \end{cases}$$

En prenant $\varepsilon > 0$ assez petit, à chaque point y_j sont associés deux zéros de $\Psi_\varepsilon(x)$ voisins de y_j . Le nombre de zéros de $\Psi_\varepsilon(x)$ dans I sera donc au moins $(p+2q)$ et, comme $\Psi_\varepsilon(x) \neq 0$, on aura :

$$p+2q \leq N-1.$$

Donc, même en comptant deux fois les zéros doubles, $\Phi(x)$ a au plus $(N-1)$ zéros dans I .

Théorème 4: soit M un sous-espace de Haar de dimension finie de $C[a,b]$ et soit $\Phi_1(x), \dots, \Phi_N(x)$ une base de M . Pour que l'origine de R^N appartienne à l'enveloppe convexe de l'ensemble de points $\{[\lambda_i \Phi_1(x_i), \dots, \lambda_i \Phi_N(x_i)] \in R^N \mid i=0,1,\dots,N\}$, où $a \leq x_0 < x_1 < \dots < x_N \leq b$ et $\lambda_i \neq 0$, il faut et il suffit que l'on ait: $\lambda_i \lambda_{i-1} < 0$ ($i=1,\dots,N$) CHENEY [1966].

Etant donnés N points $z_1, z_2, \dots, z_N$ dans I , posons:

$$(10) \quad D(z_1, \dots, z_N) = \begin{vmatrix} \Phi_1(z_1) & \Phi_1(z_2) & \dots & \Phi_1(z_N) \\ \dots & \dots & \dots & \dots \\ \Phi_N(z_1) & \Phi_N(z_2) & \dots & \Phi_N(z_N) \end{vmatrix}$$

D'après (8), lorsque les points $z_1, z_2, \dots, z_N$ sont deux à deux distincts, on a: $D(z_1, \dots, z_N) \neq 0$

De plus, si l'on a:

$$a \leq y_1 < \dots < y_N \leq b \quad \text{et} \quad a \leq z_1 < \dots < z_N \leq b,$$

alors les deux déterminants $D(y_1, \dots, y_N)$ et $D(z_1, \dots, z_N)$ ont même signe.

Sinon, on aurait par exemple:

$$D(y_1, \dots, y_N) < 0 < D(z_1, \dots, z_N)$$

D étant une application continue de R^N dans R , on pourrait trouver $\lambda \in]0,1[$ tel que:

$$D[\lambda y_1 + (1-\lambda)z_1, \dots, \lambda y_N + (1-\lambda)z_N] = 0$$

D'après (8), cela signifierait qu'il existe deux indices i et j distincts tels que:

$$\lambda y_i + (1-\lambda)z_i = \lambda y_j + (1-\lambda)z_j$$

ou:

$$\lambda(y_i - y_j) = (1-\lambda)(z_j - z_i)$$

c'est-à-dire que $(y_i - y_j)$ et $(z_i - z_j)$ seraient de signes opposés ce qui est impossible.

$D(y_1, \dots, y_N)$ et $D(z_1, \dots, z_N)$ ont donc bien même signe.

L'origine de R^N appartient à l'enveloppe convexe de l'ensemble considéré si et seulement si : il existe une solution $(\theta_0, \theta_1, \dots, \theta_N)$ avec $\theta_i \geq 0$, $\sum \theta_i \neq 0$ du système

$$\sum_{i=0}^N \theta_i \lambda_i \begin{pmatrix} \Phi_1(x_i) \\ \vdots \\ \Phi_N(x_i) \end{pmatrix} = 0$$

Ce système s'écrit encore (en choisissant $\theta_0 \neq 0$)

$$\sum_{i=1}^N -\frac{\theta_i \lambda_i}{\theta_0 \lambda_0} \Phi_j(x_i) = \Phi_j(x_0) \quad j = 1, \dots, N$$

et la solution est :

$$-\frac{\theta_i \lambda_i}{\theta_0 \lambda_0} = \frac{D(x_1, \dots, x_{i-1}, x_0, x_{i+1}, \dots, x_N)}{D(x_1, \dots, x_N)}$$

ou encore :

$$-\frac{\theta_i \lambda_i}{\theta_0 \lambda_0} = (-1)^{i-1} \frac{D(x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N)}{D(x_1, \dots, x_N)}$$

Le second membre de cette relation est du signe de $(-1)^{i-1}$ puisque les deux déterminants qui y figurent sont du même signe. Pour que la solution $(\theta_0, \theta_1, \dots, \theta_N)$ soit positive, il faut et il suffit donc que :

$$\text{sign}(\lambda_i) = (-1)^i \text{sign}(\lambda_0) \quad i = 1, \dots, N$$

Nous allons maintenant démontrer un résultat analogue au théorème 4 mais valable dans le cas où les éléments Φ de M s'annulent tous en des points ξ_i donnés.

Soient $\xi_1, \xi_2, \dots, \xi_K$ des points donnés, appelés "noeuds", contenus dans un intervalle I' contenant $[a, b]$. Désignons par $\tilde{I}$ l'ensemble

obtenu en enlevant de I les noeuds contenus dans cet intervalle.

Soit M un sous-espace de dimension finie de $C[I]$ et $\Phi_1(x)$, $\Phi_2(x), \dots, \Phi_N(x)$ une base de M . Nous dirons que M vérifie la condition de Haar sur $\tilde{I}$ si quels que soient les points $x_1, \dots, x_N$ distincts dans $\tilde{I}$ la relation (8) est vérifiée.

Nous noterons $\nu^*(M) - 1$ le nombre maximum de zéros dans l'ensemble $I \cup \{\xi_1, \dots, \xi_K\}$ des éléments $\Phi \neq 0$ de M , où l'on compte deux fois les zéros intérieurs à I où Φ ne change pas de signe.

Remarquons que si $\nu^*(M) \leq K+N$, alors tout élément Φ de M a au plus $(N-1)$ zéros dans $\tilde{I}$ et donc M vérifie la condition de Haar sur $\tilde{I}$ (la réciproque étant fausse à cause de la présence possible de noeuds comme zéros doubles).

Théorème 5: soit M un sous-espace de dimension finie de $C[a, b]$ et soit $\Phi_1(x), \dots, \Phi_N(x)$ une base de M . Supposons que tout élément $\Phi \in M$ vérifie $\Phi(\xi_i) = 0$ ($i = 1, \dots, K$) et que: $\nu^*(M) \leq K+N$. Pour que l'origine de R^N appartienne à l'enveloppe convexe de l'ensemble de points $\{[\lambda_i \Phi_1(x_i), \dots, \lambda_i \Phi_N(x_i)] \in R^N \mid i = 0, 1, \dots, N\}$ où $a \leq x_0 < x_1 < \dots < x_N \leq b$, $x_i \in \tilde{I}$ et $\lambda_i \neq 0$, il faut et il suffit que l'on ait:

$\lambda_i \lambda_{i+1} < 0$ si $]x_i, x_{i+1}[$ contient un nombre pair (ou nul) de noeuds,
 $\lambda_i \lambda_{i+1} > 0$ si $]x_i, x_{i+1}[$ contient un nombre impair de noeuds.

Etant donnés N points $z_1, z_2, \dots, z_N$ de $\tilde{I}$ tels que $a \leq z_1 < \dots < z_N \leq b$, posons

$$D(z_1, \dots, z_N) = \begin{vmatrix} \Phi_1(z_1) & \dots & \Phi_1(z_N) \\ \dots & \dots & \dots \\ \Phi_N(z_1) & \dots & \Phi_N(z_N) \end{vmatrix}$$

D'après nos hypothèses M vérifie la condition de Haar sur $\tilde{I}$ et par suite:

$$D(z_1, \dots, z_N) \neq 0$$

Pour que l'origine de R^N appartienne à l'enveloppe convexe de l'ensemble considéré, il faut et il suffit qu'il existe une solution $(\theta_0, \theta_1, \dots, \theta_N)$

avec $\theta_i \geq 0$, $\sum \theta_i \neq 0$ du système :

$$\sum_{i=0}^N \theta_i \lambda_i \begin{pmatrix} \Phi_1(x_i) \\ \vdots \\ \Phi_N(x_i) \end{pmatrix} = 0$$

Ce système s'écrit encore (en choisissant $\theta_p \neq 0$) :

$$\sum_{\substack{i=0 \\ i \neq p}}^N - \frac{\theta_i \lambda_i}{\theta_p \lambda_p} \Phi_j(x_i) = \Phi_j(x_p) \quad j = 1, \dots, N$$

et la solution est :

$$- \frac{\theta_{p+1} \lambda_{p+1}}{\theta_p \lambda_p} = \frac{D(x_0, \dots, x_{p-1}, x_p, x_{p+2}, \dots, x_N)}{D(x_0, \dots, x_{p-1}, x_{p+1}, x_{p+2}, \dots, x_N)}$$

La fonction $\Psi(x) = D(x_0, \dots, x_{p-1}, x, x_{p+2}, \dots, x_N)$ est un élément de M , non identiquement nul, qui s'annule aux points $x_0, \dots, x_{p-1}, x_{p+2}, \dots, x_N$ et $\xi_1, \dots, \xi_K$, c'est à dire en $N+K-1$ points.

Comme on a : $\nu^*(M) \leq K+N$, il en résulte que tous les noeuds intérieurs à I sont des zéros simples de $\Psi(x)$ c'est-à-dire que $\Psi(x)$ s'y annule en changeant de signe.

Donc s'il y a r noeuds dans $]x_p, x_{p+1}[$, on a :

$$(-1)^r \Psi(x_p) \Psi(x_{p+1}) > 0$$

Pour que la solution $(\theta_0, \theta_1, \dots, \theta_N)$ soit positive, il faut et il suffit donc que :

$$(-1)^{r+1} \lambda_p \lambda_{p+1} > 0$$

c'est-à-dire :

$$\lambda_p \lambda_{p+1} < 0 \quad \text{si} \quad r \text{ est pair ou nul}$$

$$\lambda_p \lambda_{p+1} > 0 \quad \text{si} \quad r \text{ est impair.}$$

Chapitre. I

APPROXIMATION RATIONNELLE AVEC CONTRAINTES GÉNÉRALES

Dans ce chapitre, on étudie l'approximation d'une fonction continue $f(x)$ par des "fractions rationnelles généralisées"

$$R(x) = \frac{\sum_{i=0}^m a_i g_i(x)}{\sum_{i=0}^n b_i h_i(x)}$$

dont les coefficients vérifient la condition :

$$(a_0, \dots, a_m, b_0, \dots, b_n) \in \Omega$$

où Ω désigne un ensemble donné dans R^{m+n+2} .

Le problème est posé dans le paragraphe 1. L'existence d'une meilleure approximation est étudiée au paragraphe 2. On montre au paragraphe 3 que l'ensemble des éléments optimaux est convexe (en un certain sens). Dans les paragraphes 4 et 5, on donne une généralisation d'un théorème de La Vallée-Poussin (permettant de calculer une borne inférieure de $\|f-R\|$) et une caractérisation de l'élément optimal. Au paragraphe 6 on montre que le point $(a_0^*, \dots, a_m^*, b_0^*, \dots, b_n^*)$ dont les coordonnées sont les coefficients de l'élément optimal appartient à la frontière de Ω . Enfin dans les paragraphes 7 et 8, on étudie l'unicité et les propriétés d'alternance. Au paragraphe 9, on aborde le cas d'une norme faisant intervenir les dérivées.

1-Introduction.

Soit X un espace compact et $C[X]$ l'espace vectoriel des fonctions continues sur X , à valeurs réelles. On se donne dans $C[X]$ des fonctions $g_0(x), g_1(x), \dots, g_m(x), h_0(x), h_1(x), \dots, h_n(x)$. On suppose les fonctions $g_0(x), g_1(x), \dots, g_m(x)$ d'une part, $h_0(x), h_1(x), \dots, h_n(x)$ d'autre part, linéairement indépendantes sur X . On se donne enfin un ensemble Ω fermé non vide dans R^{m+n+2} .

On désigne par $W_{m,n}$ la classe des fonctions:

$$R(x) = \frac{\sum_{i=0}^m a_i g_i(x)}{\sum_{i=0}^n b_i h_i(x)}$$

où les a_i, b_i sont des nombres réels vérifiant les conditions :

$$(1) \quad (a_0, a_1, \dots, a_m, b_0, b_1, \dots, b_n) \in \Omega$$

$$(2) \quad \sum_{i=0}^n b_i h_i(x) > 0 \quad \text{sur } X.$$

Problème: soit $f(x) \in C[X]$, $f \in W_{m,n}$. On cherche un élément $R^* \in W_{m,n}$ tel que :

$$\|f - R^*\| \leq \|f - R\| \quad \forall R \in W_{m,n}$$

la norme d'une fonction $g \in C[X]$ étant définie par :

$$\|g\| = \max_{x \in X} |g(x)|$$

Un tel élément R^* , s'il existe, sera dit *optimal* (pour la fonction f dans la classe $W_{m,n}$).

On étudiera l'existence d'un élément optimal, la caractérisation de cet élément, l'unicité et la continuité de l'opérateur qui à la fonction $f \in C[X]$ associe sa meilleure approximation $R^* \in W_{m,n}$.

Le cas particulier des polynômes ($n=0$) a déjà été étudié : PASZKOWSKI [1955, 1957] a étudié l'approximation d'une fonction par des polynômes égaux à la fonction en des points donnés. BROSOWSKI [1965] a étudié le cas où les coefficients du polynôme d'approximation vérifient certaines égalités linéaires. GOLSHTEIN [1961, 1961a] et TERRENOIRE [1967] ont envisagé le cas où les coefficients sont soumis à des égalités et inégalités linéaires. Enfin BUROV [1961] et RICE [1963] ont considéré des contraintes générales du type (1).

D'autre part, l'approximation par fractions rationnelles généralisées a été étudiée par de nombreux auteurs : COLLATZ [1960, 1964] et surtout BOEHM [1965] ont étudié l'existence d'une meilleure approximation. BROSOWSKI [1965a] CHENEY et LOEB [1964], CHENEY [1966] ont étudié la caractérisation de l'élément optimal. CHENEY [1964, 1966] a donné des propriétés d'alternance qui généralisent les résultats classiques de Tchebysheff. L'unicité a été spécialement étudiée par BROSOWSKI [1965b] et LOEB [1966]. La continuité de l'opérateur est traitée dans CHENEY [1964, 1966], CHENEY et LOEB [1966], WERNER [1964].

Enfin, MEINARDUS [1964] a donné une théorie de l'approximation non linéaire (approximation par des fonctions $F(a_1, \dots, a_n; x)$ dépendant non linéairement de n paramètres $a_1, \dots, a_n$). Cette théorie, appliquée à l'approximation par fractions rationnelles généralisées permet de retrouver la caractérisation des éléments optimaux et certains résultats concernant les propriétés d'alternance et l'unicité.

Remarque : Le problème considéré ici contient le cas où l'on prend comme norme sur $C[X]$ la norme suivante :

$$||g|| = \max_{x \in X} u(x) |g(x)|$$

où $u(x)$ est une fonction donnée dans $C[X]$ strictement positive sur X .

En effet, le problème de rendre minimum la quantité :

$$\max_{x \in X} u(x) \left| f(x) - \frac{\sum a_i g_i(x)}{\sum b_i h_i(x)} \right|$$

s'énonce aussi : rendre minimum la quantité :

$$\max_{x \in X} \left| f^*(x) - \frac{\sum a_i g_i^*(x)}{\sum b_i h_i(x)} \right|$$

où l'on a posé :

$$f^*(x) = u(x)f(x), \quad g_i^*(x) = u(x) g_i(x)$$

Notations :

Pour $R \in W_{m,n}$, on note E_R l'ensemble des (e) -points de $f-R$, c'est-à-dire l'ensemble des points de X tels que :

$$f(x) - R(x) = \pm ||f-R||$$

On désigne par $\mathcal{D}_R$ la classe des fonctions:

$$\Phi(x) = \sum_{i=0}^m a_i^* g_i(x) - R(x) \sum_{i=0}^n b_i^* h_i(x)$$

dont les coefficients a_i^*, b_i^* sont des nombres réels vérifiant (1) .

On pose enfin:

$$y_R(x) = f(x) - R(x)$$

$$\rho = \inf_{R \in W_{m,n}} \|f - R\|$$

2-Existence d'un élément optimal.

BOEHM [1965] a démontré l'existence d'une solution du problème en supposant que les coefficients du dénominateur vérifient la relation:

$$|b_0| + |b_1| + \dots + |b_n| = 1.$$

Nous allons reprendre la même démonstration en faisant des hypothèses nous permettant de nous ramener à l'un des deux cas suivants :

$$\sum_{i=0}^n |b_i| = 1 \quad \text{ou:} \quad 0 < \sum_{i=0}^n |b_i| \leq B$$

Désignons par $\bar{W}_{m,n}$ la classe des fonctions $R \in C[X]$ qui vérifient une identité de la forme :

$$(3) \quad R(x) \sum_{i=0}^n b_i h_i(x) = \sum_{i=0}^m a_i g_i(x)$$

où les a_i, b_i sont des nombres réels non tous nuls vérifiant (1).

Remarquons que:

$$W_{m,n} \subseteq \bar{W}_{m,n}$$

Théorème 1: Si les fonctions $h_0(x), h_1(x), \dots, h_n(x)$ sont linéairement indépendantes sur tout sous-ensemble ouvert de X et si Ω vérifie l'une des deux hypothèses suivantes:

(H1) Il existe un nombre B tel que tout point $(a_0, \dots, a_m, b_0, \dots, b_n) \in \Omega$ vérifie: $B \geq \sum |b_j| > 0$

(H2) Pour tout $\lambda > 0$, l'ensemble $\lambda\Omega$ est contenu dans Ω , alors toute fonction $f \in C[X]$ a dans $\bar{W}_{m,n}$ au moins un élément optimal.

Soit $R_1(x), R_2(x), \dots, R_k(x), \dots$, une suite d'éléments de $\bar{W}_{m,n}$ tels que:

$$\|f - R_k\| \rightarrow \bar{\rho} = \inf_{R \in \bar{W}_{m,n}} \|f - R\|$$

D'après (3), on peut écrire:

$$R_k(x) Q_k(x) \equiv P_k(x)$$

avec:

$$P_k(x) = \sum_{j=0}^m a_j^{(k)} g_j(x), \quad Q_k(x) = \sum_{j=0}^n b_j^{(k)} h_j(x)$$

et:

$$(a_0^{(k)}, \dots, a_m^{(k)}, b_0^{(k)}, \dots, b_n^{(k)}) \in \Omega$$

Dans le cas où l'hypothèse (H1) est vérifiée, on a:

$$\sum_{j=0}^n |b_j^{(k)}| \leq B$$

et dans le cas où l'hypothèse (H2) est vérifiée, on peut supposer:

$$\sum_{j=0}^n |b_j^{(k)}| = B' \neq 0$$

Dans un cas comme dans l'autre, on en déduit:

$$\|Q_k\| \leq B \max_j \|h_j\| = M$$

D'autre part, à partir d'un certain rang, on aura:

$$\|\mathcal{F} - R_k\| \leq \bar{\rho} + 1$$

D'où:

$$(4) \quad \|P_k\| \leq \|Q_k\| \cdot \|R_k\| \leq M(\|\mathcal{F}\| + \bar{\rho} + 1)$$

Les fonctions $g_0(x), g_1(x), \dots, g_m(x)$ étant linéairement indépendantes sur X , la relation (4) montre qu'il existe une constante A indépendante de k telle que:

$$\sum_{j=0}^m |a_j^{(k)}| \leq A$$

A chaque fraction $R_k(x)$, nous associons le point M_k de coordonnées $(a_0^{(k)}, \dots, a_m^{(k)}, b_0^{(k)}, \dots, b_n^{(k)})$. Tous les points M_k sont dans une partie fermée bornée de R^{m+n+2} , d'après ce que nous venons de voir. Il existe donc une sous-suite de points M_k qui converge vers un point M , c'est-à-dire que l'on peut extraire de la suite initiale de fonctions $R_1(x), R_2(x), \dots, R_k(x)$ une sous-suite (que nous noterons encore $R_k(x)$) dont les coefficients ont une limite:

$$P_k(x) \rightarrow P(x) = \sum_{j=0}^m a_j g_j(x)$$

$$Q_k(x) \rightarrow Q(x) = \sum_{j=0}^n b_j h_j(x)$$

Les points $M_1, M_2, \dots, M_k, \dots$ sont dans Ω . Il en est donc de même de leur limite M , c'est-à-dire que les coefficients de $P(x)$ et $Q(x)$ vérifient :

$$(a_0, \dots, a_m, b_0, \dots, b_n) \in \Omega$$

De plus, on a:

$$\sum_{j=0}^m |b_j| > 0$$

(car cette somme est strictement positive lorsque (H1) est vérifiée et égale à B' lorsque (H2) est vérifiée).

Il reste à montrer qu'il existe une fonction $R \in C[X]$ telle que:

$$Q(x)R(x) = P(x)$$

et à vérifier que pour cette fonction, on a:

$$\|f - R\| = \bar{\rho}$$

D'après l'indépendance linéaire des fonctions $h_0(x), h_1(x), \dots, h_n(x)$ sur tout sous-ensemble ouvert de X , la fonction $Q(x)$ est différente de zéro sur un ensemble Y dense dans X .

Nous allons poser:

$$R(x) = \frac{P(x)}{Q(x)} \quad \text{si } x \in Y$$

$$R(x) = \lim_{x \in Y, x \rightarrow x_0} \frac{P(x)}{Q(x)} \quad \text{si } x \in X - Y$$

On définit ainsi une fonction $R \in C[X]$ et donc $R \in \bar{W}_{m,n}$. Calculons maintenant $\|f-R\|$:

Si $x \in Y$, on a :

$$\begin{aligned} |f(x) - R(x)| &\leq |f(x) - R_k(x)| + |R(x) - R_k(x)| \\ &\leq \|f - R_k\| + |R(x) - R_k(x)| \end{aligned}$$

D'où, en passant à la limite pour $k \rightarrow \infty$:

$$(6) \quad |f(x) - R(x)| \leq \bar{\rho} \quad \forall x \in Y$$

Si $x \in X - Y$, prenons une suite de points $x_i \in Y$ tels que :

$$(7) \quad |R(x) - R(x_i)| < \varepsilon_i \quad \text{et} \quad |f(x) - f(x_i)| < \varepsilon_i$$

avec :

$$\lim_{i \rightarrow \infty} \varepsilon_i = 0$$

On a alors :

$$|f(x) - R(x)| \leq |f(x) - f(x_i)| + |f(x_i) - R(x_i)| + |R(x_i) - R(x)|$$

Soit, d'après (6) et (7) :

$$|f(x) - R(x)| \leq \bar{\rho} + 2 \varepsilon_i$$

Et, en passant à la limite :

$$|f(x) - R(x)| \leq \bar{\rho} \quad \forall x \in X - Y$$

Donc finalement :

$$\|f - R\| \leq \bar{\rho}$$

ce qui compte tenu de l'inégalité évidente $\|f - R\| \geq \bar{\rho}$ prouve que R est élément optimal pour f dans la classe $\bar{W}_{m,n}$.

Remarque 1 : Prenons pour X un intervalle $[a,b]$ de la droite réelle et supposons les fonctions $g_0(x), g_1(x), \dots, h_n(x)$ analytiques sur $[a,b]$. Ces fonctions sont indépendantes sur tout intervalle contenu dans $[a,b]$. Le théorème 1 est dans ce cas particulier une généralisation du résultat donné par CHENEY [1966] lorsqu'il n'y a aucune contrainte sur les coefficients.

Remarque 2 : Dans certains cas particuliers, on pourra montrer qu'il existe un élément optimal pour f dans $W_{m,n}$ (au lieu de $\bar{W}_{m,n}$).

Par exemple, si pour tout point $(a_0, \dots, a_m, b_0, \dots, b_n) \in \Omega$, on a l'inégalité :

$$\sum_{j=0}^n b_j h_j(x) \geq \varepsilon > 0 \quad \forall x \in X$$

les ensembles $W_{m,n}$ et $\bar{W}_{m,n}$ sont identiques.

Un autre exemple peut s'obtenir en prenant pour X un intervalle $[a, b]$ de la droite réelle et pour classe $W_{m,n}$ l'ensemble des fractions rationnelles "ordinaires" :

$$R(x) = \frac{a_0 + a_1 x + \dots + a_m x^m}{b_0 + b_1 x + \dots + b_n x^n}$$

telles que :

$$g(x) \leq R(x) \leq h(x) \quad \forall x \in [a, b]$$

$$b_0 + b_1 x + \dots + b_n x^n > 0 \quad \forall x \in [a, b]$$

où $g(x)$ et $h(x)$ sont deux fonctions données dans $C[a, b]$.

D'une part, l'hypothèse (H2) est vérifiée et d'autre part, on peut montrer que tout élément de $\bar{W}_{m,n}$ qui est limite d'une suite d'éléments de $W_{m,n}$ peut (après division éventuelle du "numérateur" et du "dénominateur" par un facteur commun) être identifié avec un élément de $W_{m,n}$.

Ainsi dans ce cas, toute fonction $f \in C[a, b]$ a dans $W_{m,n}$ au moins un élément optimal.

Remarque 3 : Lorsqu'on abandonne les hypothèses (H1) et (H2), on peut rencontrer des cas de non existence (qui correspondent à des coefficients qui tendent vers l'infini).

Par exemple, prenons $X = [0, 1]$ et $f(x) \equiv 2x$

Soit $W_{m,n}$ la classe des fractions :

$$R(x) = \frac{a_0 + a_1 x + \dots + a_m x^m}{b_0 + b_1 x + \dots + b_n x^n}$$

telles que :

$$a_0 = b_0 + 1, \quad b_0 \geq 1$$

Pour tout $R \in W_{m,n}$, on a :

$$\|f - R\| \geq |f(0) - R(0)| = 1 + \frac{1}{b_0} > 1$$

D'autre part, considérons dans $W_{m,n}$ la fonction :

$$R_p(x) = \frac{(p+1) + x}{p+x} = 1 + \frac{1}{x+p} \quad (p \geq 1)$$

On a :

$$\|f - R_p\| = 1 + \frac{1}{p}$$

ce qui prouve que :

$$\rho = \inf_{W_{m,n}} \|f - R\| = 1$$

Comme on a vu que : $\|f - R\| > 1$ pour tout $R \in W_{m,n}$, il n'y a pas d'élément optimal pour f dans $W_{m,n}$.

Si on considère une fonction $R \in \bar{W}_{m,n}$, on a :

$$R(x) [b_0 + b_1 x + \dots + b_n x^n] = [a_0 + a_1 x + \dots + a_m x^n]$$

D'où :

$$R(0) = 1 + \frac{1}{b_0}$$

ce qui montre que l'on a aussi :

$$\bar{\rho} = 1 \quad \text{et} \quad \|f - R\| > 1 \quad \forall R \in \bar{W}_{m,n}$$

f n'a donc pas non plus d'élément optimal dans $\bar{W}_{m,n}$.

Il faut remarquer que lorsque $p \rightarrow \infty$, la fonction $R_p(x)$ converge uniformément vers 1 qui vérifie $\|f - 1\| = 1$ mais n'appartient pas à $W_{m,n}$.

3 - Convexité de l'ensemble des éléments optimaux.

Théorème 2 : supposons Ω convexe. Si $R_0(x) = P_0(x)/Q_0(x)$ et

$R_1(x) = P_1(x)/Q_1(x)$ sont optimaux pour f dans $W_{m,n}$, il en est de même pour :

$$R_\alpha(x) = \frac{\alpha P_1(x) + (1-\alpha)P_0(x)}{\alpha Q_1(x) + (1-\alpha)Q_0(x)} \quad \text{où } 0 \leq \alpha \leq 1$$

Cela revient à dire que l'ensemble des points de R^{m+n+2} dont les coordonnées $(a_0, a_1, \dots, a_m, b_0, b_1, \dots, b_n)$ sont les coefficients des éléments optimaux est convexe.

Tout d'abord, $R_\alpha \in W_{m,n}$ car Ω est convexe. La propriété énoncée se déduit immédiatement de l'égalité suivante :

$$(8) \quad f(x) - R_\alpha(x) = \frac{\alpha Q_1(x)}{\alpha Q_1(x) + (1-\alpha)Q_0(x)} \left[f(x) - \frac{P_1(x)}{Q_1(x)} \right] + \frac{(1-\alpha)Q_0(x)}{\alpha Q_1(x) + (1-\alpha)Q_0(x)} \left[f(x) - \frac{P_0(x)}{Q_0(x)} \right]$$

La formule (8) montre en outre que si $x_0 \in X$ est un (e)-point de $f - R_\alpha$, c'est aussi un (e)-point de même nature pour $f - R_0$ et $f - R_1$ c'est-à-dire que l'on a :

$$f(x) - R_0(x) = f(x) - R_1(x) = f(x) - R_\alpha(x) \quad \forall x \in E_{R_\alpha}$$

ou encore :

$$(9) \quad R_0(x) = R_1(x) = R_\alpha(x) \quad \forall x \in E_{R_\alpha}$$

4 - Extension du théorème de la Vallée - Poussin.

L'application des résultats de MEINARDUS [1964, th.80] au cas de l'approximation rationnelle avec contraintes nous donne le résultat suivant :

Soit $R(x) \in W_{m,n}$. S'il existe un ensemble $D \subset X$ tel que :

$$a) \quad f(x) - R(x) \neq 0 \quad \forall x \in D,$$

$$b) \quad \inf_{x \in D} [R_1(x) - R(x)] [f(x) - R(x)] \leq 0 \quad \forall R_1 \in W_{m,n},$$

alors, on a :

$$\rho \geq \inf_{x \in D} |f(x) - R(x)|$$

Nous allons démontrer une forme légèrement différente de ce résultat (forme qui nous sera utile lorsque nous expliciterons la définition de Ω à l'aide d'égalités et inégalités linéaires).

Théorème 3 : Soit $R \in W_{m,n}$. S'il existe un ensemble $D \subset X$ tel que :

$$a) \quad f(x) - R(x) \neq 0 \quad \forall x \in D,$$

$$b) \quad \inf_{x \in D} \Phi(x) [f(x) - R(x)] \leq 0 \quad \forall \Phi \in \mathcal{D}_R;$$

alors, on a :

$$\rho \geq \inf_{x \in D} |f(x) - R(x)|$$

Supposons au contraire que l'on ait :

$$(10) \quad \rho < \inf_{x \in D} |f(x) - R(x)|$$

Alors, il existe dans $W_{m,n}$ un élément $R_1(x) = \frac{P_1(x)}{Q_1(x)}$ tel que :

$$\|f - R_1\| < \inf_{x \in D} |f(x) - R(x)|$$

et donc :

$$(11) \quad |f(x) - R_1(x)| < |f(x) - R(x)| \quad \forall x \in D$$

Considérons dans $\mathcal{D}_R$ l'élément suivant :

$$\Phi(x) = P_1(x) - R(x) Q_1(x)$$

qu'on peut encore écrire :

$$(12) \quad \Phi(x) = Q_1(x)[R_1(x) - R(x)] = Q_1(x)\{[f(x) - R(x)] - [f(x) - R_1(x)]\}$$

De (11) et (12), on déduit immédiatement :

$$\Phi(x) [f(x) - R(x)] > 0 \quad \forall x \in D$$

ce qui est en contradiction avec l'hypothèse b).

La relation (10) est donc fausse et l'on a bien :

$$\rho \geq \inf_{x \in D} |f(x) - R(x)|$$

Remarque 1 : on utilise l'hypothèse b) seulement pour les fonctions

$\Phi(x) = P_1(x) - R(x)Q_1(x)$ de $\mathcal{D}_R$ telles que $Q(x) > 0$ sur X . En se restreignant dans b) à de telles fonctions, on retrouve le résultat cité au début de ce paragraphe.

Remarque 2 : Du théorème 3, on peut déduire le résultat suivant :

Corollaire : soit $R \in W_{m,n}$. S'il existe des points $x_1, \dots, x_r$ de X et des nombres $\lambda_1, \dots, \lambda_r$ tels que :

$$\lambda_i [f(x_i) - R(x_i)] > 0 \quad i = 1, \dots, r$$

$$\sum_{i=1}^r \lambda_i \Phi(x_i) \leq 0 \quad \forall \Phi \in \mathcal{D}_R$$

alors :

$$\rho \geq \min_{1 \leq i \leq r} |f(x_i) - R(x_i)|$$

En effet, les formules ci-dessus montrent qu'il n'existe aucun élément $\Phi \in \mathcal{D}_R$ tel que :

$$\Phi(x_i) [f(x_i) - R(x_i)] > 0 \quad (i = 1, \dots, r)$$

En posant $D = \{x_1, \dots, x_r\}$, on a donc :

$$\inf_{x \in D} \Phi(x) [f(x) - R(x)] \leq 0 \quad \forall \Phi \in \mathcal{D}_R$$

et on obtient le résultat en appliquant le théorème 3.

5 - Caractérisation des éléments optimaux.

De la théorie de l'approximation non linéaire (MEINARDUS [1964]), on déduit le résultat suivant en supposant Ω convexe : pour que $R \in W_{m,n}$ soit optimal, il faut et il suffit que l'on ait :

$$\inf_{x \in E_R} [R_1(x) - R(x)] [f(x) - R(x)] \leq 0 \quad \forall R_1 \in W_{m,n}$$

Comme au paragraphe précédent, nous allons démontrer une forme légèrement différente de ce résultat, analogue au théorème donné par CHENEY [1966] dans le cas de l'approximation par fractions rationnelles généralisées sans contraintes sur les coefficients. Le théorème ci-dessous généralise celui donné par BJUROV [1961] dans le cas de l'approximation polynomiale avec contraintes non linéaires.

Théorème 4 : pour que $R \in W_{m,n}$ soit optimal pour f , il est suffisant - et, dans le cas où Ω est convexe, il est nécessaire - que l'on ait :

$$\inf_{x \in E_R} \Phi(x) [f(x) - R(x)] \leq 0 \quad \forall \Phi \in \mathcal{D}_R$$

La condition est évidemment suffisante d'après le théorème 3.

Supposons Ω convexe et montrons qu'elle est nécessaire. Remarquons tout d'abord que E_R est compact car c'est une partie fermée de X compact.

Soit $R(x)$ un élément de $W_{m,n}$ et supposons qu'il existe dans $\mathcal{D}_R$ un élément :

$$\Phi(x) = P_0(x) - R(x)Q_0(x)$$

tel que :

$$(13) \quad \Phi(x)[f(x) - R(x)] > 0 \quad \forall x \in E_R$$

Nous allons montrer que $R(x)$ n'est pas optimal.

Posons :

$$R(x) = \frac{P(x)}{Q(x)}, \quad R_1(x) = \frac{(1-\lambda)P(x) + \lambda P_0(x)}{(1-\lambda)Q(x) + \lambda Q_0(x)}$$

Pour $0 < \lambda < 1$, les coefficients de $R_1(x)$ vérifient, d'après la convexité de Ω , la condition :

$$(a_0, a_1, \dots, a_m, b_0, b_1, \dots, b_n) \in \Omega$$

De plus, puisque $Q(x) > 0$ sur X , pour $\lambda > 0$ assez petit, on aura :

$$(1-\lambda)Q(x) + \lambda Q_0(x) > 0 \quad \text{sur } X,$$

c'est-à-dire que $R_1(x)$ appartient à $W_{m,n}$.

Nous allons maintenant montrer qu'on peut choisir λ de manière à ce que :

$$\|f - R_1\| < \|f - R\|$$

On a :

$$\begin{aligned} (14) \quad f(x) - R_1(x) &= [f(x) - R(x)] + [R(x) - R_1(x)] \\ &= y_R(x) - \frac{\lambda \Phi(x)}{(1-\lambda)Q(x) + \lambda Q_0(x)} \end{aligned}$$

Posons :

$$h(x) = \text{sign } y_R(x)$$

$$\delta = \inf_{x \in E_R} h(x) \Phi(x)$$

On a $\delta > 0$ [sinon on pourrait trouver une suite de (e)-points qui convergerait vers un (e)-point x_0 tel que $\Phi(x_0) y(x_0) = 0$, ce qui est en contradiction avec (13)] .

Considérons les deux ensembles :

$$F_1 = \{x \in X \mid h(x) \Phi(x) > \frac{\delta}{2} \text{ et } \|y(x)\| > \frac{1}{2} \|f - R\|\}$$

$$F_2 = X - F_1$$

F_2 est un compact ne contenant aucun (e)-point. Donc il existe un nombre $\mu > 0$ tel que :

$$\|f(x) - R(x)\| \leq \mu < \|f - R\| \quad \forall x \in F_2$$

Pour $x \in F_2$, on aura donc, d'après (14) :

$$\|f(x) - R_1(x)\| \leq \mu + \|R(x) - R_1(x)\|$$

Or lorsque $\lambda \rightarrow 0$, $\|R(x) - R_1(x)\| \rightarrow 0$ et donc en choisissant $\lambda > 0$ assez petit, on aura :

$$(15) \quad \|f(x) - R_1(x)\| < \|f - R\| \quad \forall x \in F_2$$

D'autre part, on peut choisir λ assez petit pour que :

$$\|R(x) - R_1(x)\| < \frac{1}{2} \|f - R\|$$

et alors $f(x) - R(x)$ et $f(x) - R_1(x)$ ont même signe sur F_1 .

Pour $x \in F_1$, on peut alors écrire :

$$\|f(x) - R_1(x)\| = h(x) [f(x) - R_1(x)]$$

$$= h(x) y_R(x) - \frac{\lambda h(x) \Phi(x)}{(1 - \lambda) Q(x) + \lambda Q_0(x)}$$

D'où l'on déduit :

$$(16) \quad \|f(x) - R_1(x)\| \leq \|f - R\| - \frac{\lambda \delta}{2 \|(1 - \lambda) Q + \lambda Q_0\|} \quad \forall x \in F_1$$

Les relations (15) et (16) montrent qu'en choisissant $\lambda > 0$ assez petit, on aura :

$$\|f - R_1\| < \|f - R\|$$

ce qui prouve que R n'est pas optimal.

Corollaire : pour que $R \in W_{m,n}$ soit optimal, il suffit qu'il existe des (e)-points $x_1, \dots, x_r$ ($r \geq 1$) de $f - R$ et des nombres $\lambda_1, \dots, \lambda_r$ tels que :

$$(17) \quad \begin{aligned} \lambda_i [f(x_i) - R(x_i)] &> 0 & i = 1, \dots, r \\ \sum_{i=1}^r \lambda_i \Phi(x_i) &\leq 0 & \forall \Phi \in \mathcal{D}_R \end{aligned}$$

En effet, les conditions (17) montrent immédiatement qu'il n'existe aucun élément $\Phi \in \mathcal{D}_R$ tel que :

$$\Phi(x) [f(x) - R(x)] > 0 \quad \forall x \in E_R,$$

ce qui prouve que R est optimal.

Nous verrons au chapitre suivant que lorsque Ω est défini par des égalités et inégalités linéaires entre les coefficients a_j, b_j , la condition (17) est également nécessaire.

6 - Propriétés des coefficients de l'élément optimal.

Désignons par $\mathcal{R}_{m,n}$ la classe des fonctions :

$$R(x) = \frac{P(x)}{Q(x)} = \frac{\sum_0^m a_j g_j(x)}{\sum_0^n b_j h_j(x)}$$

telles que :

$$\sum_0^n b_j h_j(x) > 0 \quad \text{sur } X$$

On a évidemment : $W_{m,n} \subseteq \mathcal{R}_{m,n}$

Lemme : soient $R(x) = \frac{P(x)}{Q(x)}$ et $R'(x) = \frac{P'(x)}{Q'(x)}$ deux éléments de $\mathcal{R}_{m,n}$ tels que :

$$\|f - R\| < \|f - R'\|$$

On a alors :

$$\left\| f - \frac{\lambda P + (1-\lambda)P'}{\lambda Q + (1-\lambda)Q'} \right\| < \|f - R'\| \quad \forall \lambda \in]0,1[$$

On peut en effet écrire :

$$\begin{aligned} f(x) - \frac{\lambda P(x) + (1-\lambda)P'(x)}{\lambda Q(x) + (1-\lambda)Q'(x)} &= \frac{\lambda Q(x)}{\lambda Q(x) + (1-\lambda)Q'(x)} [f(x) - R(x)] \\ &+ \frac{(1-\lambda)Q'(x)}{\lambda Q(x) + (1-\lambda)Q'(x)} [f(x) - R'(x)] \end{aligned}$$

D'où :

$$\left\| f - \frac{\lambda P + (1-\lambda)P'}{\lambda Q + (1-\lambda)Q'} \right\| \leq \left\| \frac{\lambda \|f-R\| Q(x) + (1-\lambda) \|f-R'\| Q'(x)}{\lambda Q(x) + (1-\lambda)Q'(x)} \right\|$$

Et, en tenant compte de l'hypothèse $\|f - R\| < \|f - R'\|$

$$\left\| f - \frac{\lambda P + (1-\lambda)P'}{\lambda Q + (1-\lambda)Q'} \right\| < \|f - R'\| \quad \forall \lambda \in]0,1[$$

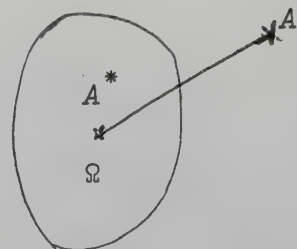
Théorème 5 : soit $R^*(x) = \sum a_j^* g_j(x) / \sum b_j^* h_j(x)$ optimal pour f dans $\mathcal{W}_{m,n}$. Si Ω est convexe et si R^* n'est pas optimal pour f dans $\mathcal{R}_{m,n}$, alors le point $(a_0^*, a_1^*, \dots, a_m^*, b_0^*, b_1^*, \dots, b_n^*)$ appartient à la frontière de Ω .

Ce résultat généralise celui donné par RICE [1963] dans le cas de l'approximation polynomiale avec contraintes.

Supposons le point A^* de coordonnées $(a_0^*, \dots, a_m^*, b_0^*, \dots, b_n^*)$ intérieur à Ω .

R^* n'étant pas optimal pour f dans $\mathcal{R}_{m,n}$, il existe dans $\mathcal{R}_{m,n}$ un élément :

$$R(x) = \frac{\sum a_j g_j(x)}{\sum b_j h_j(x)},$$



n'appartenant pas à $W_{m,n}$ et tel que :

$$\|f - R\| < \|f - R^*\|$$

Considérons dans $R_{m,n}$ la fonction suivante :

$$R_\lambda(x) = \frac{\sum_0^m [\lambda a_j + (1-\lambda)a_j^*] g_j(x)}{\sum_0^n [\lambda b_j + (1-\lambda)b_j^*] h_j(x)} \quad \text{où} \quad \lambda \in]0,1]$$

D'après le lemme précédent, on a :

$$(18) \quad \|f - R_\lambda\| < \|f - R^*\|$$

Mais pour λ assez petit, R_λ est un élément de $W_{m,n}$ (puisque le point A^* est intérieur à Ω). La formule (18) prouverait alors que R^* n'est pas optimal pour f dans $W_{m,n}$.

Ce qui montre que le point A^* appartient bien à la frontière de Ω .

Remarque : Le théorème 5 n'est intéressant que si Ω possède (relativement à R^{m+n+2}) un intérieur non vide, c'est-à-dire si Ω est de dimension $m+n+2$ (on peut trouver, dans Ω , $(m+n+3)$ points $M_1, M_2, \dots, M_N$ tels que les $(m+n+2)$ vecteurs $\vec{M_1 M_2}, \dots, \vec{M_1 M_N}$ soient linéairement indépendants). Dans le cas contraire, il existe une variété linéaire V' de dimension minimum $s < m+n+2$ uniquement déterminée qui contient Ω . De plus, l'intérieur de Ω (relativement à V') est non vide (EGGLESTON [1958], théorème 4).

Désignons par $W'_{m,n}$ l'ensemble des éléments de $R_{m,n}$ tels que :
 $(a_0, \dots, a_m, b_0, \dots, b_n) \in V'$.

On a bien entendu : $W_{m,n} \subseteq W'_{m,n} \subseteq R_{m,n}$

On a alors le résultat suivant, qui se démontre exactement comme le théorème 5 :

Corollaire : soit $R^*(x) = \sum a_j^* g_j(x) / \sum b_j^* h_j(x)$ optimal pour f dans $W_{m,n}$. Si R^* n'est pas optimal pour f dans $W'_{m,n}$, alors le point $(a_0^*, a_1^*, \dots, a_m^*, b_0^*, b_1^*, \dots, b_n^*)$ appartient à la frontière de Ω (relativement à V').

7 - Unicité.

Théorème 6 : soit R^* optimal pour f dans $W_{m,n}$. Si Ω est strictement convexe et si R^* n'est pas optimal pour f dans $R_{m,n}$, alors R^* est le seul élément optimal pour f dans $W_{m,n}$.

Ce résultat généralise celui donné par RICE [1963] dans le cas de l'approximation polynomiale avec contraintes.

Supposons qu'il existe dans $W_{m,n}$ un élément $R(x)$, différent de $R^*(x)$, optimal pour f .

Posons :

$$R(x) = \frac{\sum a_j g_j(x)}{\sum b_j h_j(x)}, \quad R^*(x) = \frac{\sum a_j^* g_j(x)}{\sum b_j^* h_j(x)}$$

D'après le théorème 2, un autre élément optimal pour f est :

$$R_\lambda(x) = \frac{\sum [\lambda a_j + (1-\lambda)a_j^*]g_j(x)}{\sum [\lambda b_j + (1-\lambda)b_j^*]h_j(x)} \quad \text{avec} \quad 0 < \lambda < 1$$

En appliquant le théorème 5, on en conclut que les points suivants appartiennent à la frontière de Ω .

$$A^* \quad (a_0^*, \dots, a_m^*, b_0^*, \dots, b_n^*)$$

$$A \quad (a_0, \dots, a_m, b_0, \dots, b_n)$$

$$A_\lambda \quad (\lambda a_0 + (1-\lambda)a_0^*, \dots, \lambda a_m + (1-\lambda)a_m^*, \lambda b_0 + (1-\lambda)b_0^*, \dots, \lambda b_n + (1-\lambda)b_n^*)$$

Ceci est en contradiction avec l'hypothèse Ω strictement convexe. Donc $R(x)$ ne peut pas exister et R^* est le seul élément optimal pour f dans $W_{m,n}$.

Remarque : comme au paragraphe précédent, le théorème 6 n'est intéressant que si Ω est de dimension $m + n + 2$.

Dans le cas contraire, introduisons la variété linéaire V' et la classe de fonctions $W'_{m,n}$ déjà considérées dans le paragraphe précédent.

Une démonstration analogue à celle du théorème 6 conduit alors au résultat suivant :

Corollaire : soit R^* optimal pour f dans $W_{m,n}$. Si Ω est strictement convexe (dans V') et si R^* n'est pas optimal pour f dans $W'_{m,n}$, alors R^* est le seul élément optimal pour f dans $W_{m,n}$.

8 - Propriétés d'alternance.

Dans ce paragraphe, nous supposons que $X = [a, b]$ intervalle fini de la droite réelle.

M étant un sous-ensemble de $C[a, b]$, nous poserons :

$\nu(M) = 1 +$ nombre maximum de changements de signe dans I des éléments de M .

Nous dirons que $f - R$ possède $N(\epsilon)$ -points alternés s'il existe N points $x_1 < \dots < x_N$ de $[a, b]$ tels que :

$$f(x_i) - R(x_i) = (-1)^i \epsilon \|f - R\| \quad \text{avec} \quad \epsilon = \pm 1$$

On a tout d'abord une nouvelle forme du théorème de la vallée Poussin (théorème 3) :

Théorème 7 : soit $R \in W_{m,n}$. S'il existe des points $x_1 < \dots < x_N$ de $[a, b]$ tels que :

$$[f(x_{i+1}) - R(x_{i+1})][f(x_i) - R(x_i)] < 0 \quad i = 1, \dots, N-1$$

$$N \geq 1 + \nu(\mathcal{D}_R)$$

alors, on a :

$$\rho \geq \min_{1 \leq i \leq N} |f(x_i) - R(x_i)|$$

D'après le théorème 3, il suffit de vérifier qu'il existe aucun élément $\Phi \in \mathcal{D}_R$ tel que :

$$\Phi(x_i) \Psi_R(x_i) > 0 \quad i = 1, \dots, N$$

Si un tel élément Φ existait, on aurait :

$$\Phi(x_i) \Phi(x_{i+1}) < 0 \quad i = 1, \dots, N-1$$

Φ présenterait donc au moins $(N - 1)$ changements de signe dans $[a, b]$ ce qui est contraire à l'hypothèse : $N - 1 \geq \nu(\mathcal{D}_R)$.

Donc Φ n'existe pas et d'après le théorème 3, on a :

$$\rho \geq \min_{1 \leq i \leq N} |f(x_i) - R(x_i)|$$

On en déduit immédiatement une condition suffisante pour que R soit optimal :

Théorème 8 : si $f - R$ possède au moins $[1 + \nu(\mathcal{D}_R)]$ (e)-points alternés, alors R est optimal pour f .

9 - Cas d'une norme faisant intervenir les dérivées.

Soit $I = [a, b]$ un intervalle compact de R . Désignons par $C_p[I]$ l'espace des fonctions réelles définies sur I , continues ainsi que leurs p premières dérivées. On munit $C_p[I]$ de la norme suivante :

$$(19) \quad \|g\| = K_0 \max_{x \in I} |g(x)| + \sum_{i=1}^p K_i \max_{x \in I} |g^{(i)}(x)|$$

avec : $K_i \geq 0$ et $\sum_{i=0}^p K_i > 0$

On se donne dans $C_p[I]$ $(m + n + 2)$ fonctions $g_0(x), g_1(x), \dots, g_m(x), h_0(x), h_1(x), \dots, h_n(x)$ et on cherche à approcher une fonction $f \in C_p[I]$ par les fonctions de la classe $W_{m,n}$, ensemble des fonctions :

$$R(x) = \frac{\sum_0^m a_i g_i(x)}{\sum_0^n b_i h_i(x)}$$

telles que :

$$(a_0, \dots, a_m, b_0, \dots, b_n) \in \Omega \text{ ensemble donné dans } R^{m+n+2}$$

$$\sum b_i h_i(x) > 0 \text{ sur } I$$

On appellera encore élément optimal pour f dans $W_{m,n}$ tout élément $R^* \in W_{m,n}$ tel que :

$$\|f - R^*\| \leq \|f - R\|$$

$$\forall R \in W_{m,n}$$

On a tout d'abord un résultat sur l'existence d'un élément optimal :

Si les fonctions $h_0(x), h_1(x), \dots, h_n(x)$ sont linéairement indépendantes sur tout sous intervalle de I et si Ω vérifie l'une des deux hypothèses suivantes :

(H1) Ω est borné et ne contient pas l'origine,

(H2) Pour tout $\lambda > 0$, l'ensemble $\lambda\Omega$ est contenu dans Ω , alors toute fonction $f \in C_p[I]$ a dans $W_{m,n}$ au moins un élément optimal.

L'ensemble $\bar{W}_{m,n}$ est défini comme la classe des fonctions $R \in C_p[I]$ qui vérifient une identité de la forme :

$$R(x) \sum_0^n b_t h_t(x) \equiv \sum_0^m a_t g_t(x)$$

où les a_t, b_t sont des nombres réels non tous nuls tels que :

$$(a_0, \dots, a_m, b_0, \dots, b_n) \in \Omega$$

La démonstration commence comme celle du théorème 1. On obtient donc une suite $R_1(x), \dots, R_k(x), \dots$ d'éléments de $\bar{W}_{m,n}$ tels que :

$$\left\{ \begin{array}{l} Q_k(x) R_k(x) \equiv P_k(x) \\ \lim_{k \rightarrow \infty} \| f - R \| = \bar{\rho} = \inf_{R \in \bar{W}_{m,n}} \| f - R \| \\ \lim_{k \rightarrow \infty} P_k(x) = P(x) = \sum_0^m a_t g_t(x) \\ \lim_{k \rightarrow \infty} Q_k(x) = Q(x) = \sum_0^n b_t h_t(x) \\ (a_0, \dots, a_m, b_0, \dots, b_n) \in \Omega \end{array} \right.$$

De plus, la formule (4) montre que l'on a :

$$\| P \| \leq (\| f \| + \bar{\rho} + 1) \| Q \|$$

Ceci prouve dans le cas où l'hypothèse (H1) est vérifiée que l'on a $\sum |b_t| \neq 0$ (sinon on aurait $a_0 = \dots = a_m = 0$ et $b_0 = \dots = b_n = 0$, ce qui est exclu). Dans le cas où c'est l'hypothèse (H2) qui est vérifiée, on a vu que l'on a

alors $\sum |b_t| = B' \neq 0$.

On a donc, dans les deux cas, $\sum |b_i| \neq 0$ c'est à dire :

$$Q(x) \neq 0$$

D'après l'hypothèse sur les fonctions $h_0(x), h_1(x), \dots, h_n(x)$ la fonction $Q(x)$ est différente de zéro sur un ensemble Y dense dans I . Nous poserons :

$$R(x) = \frac{P(x)}{Q(x)} \quad \text{si } x \in Y$$

$$R(x_0) = \lim_{x \in Y, x \rightarrow x_0} \sup \frac{P(x)}{Q(x)} \quad \text{si } x_0 \in I - Y$$

et, pour $i = 1, \dots, p$:

$$R^{(i)}(x) = \frac{d^i}{dx^i} \left[\frac{P(x)}{Q(x)} \right] \quad \text{si } x \in Y$$

$$R^{(i)}(x_0) = \lim_{x \in Y, x \rightarrow x_0} \sup \frac{d^i}{dx^i} \left[\frac{P(x)}{Q(x)} \right] \quad \text{si } x_0 \in I - Y$$

On définit ainsi une fonction $R \in C_p[I]$ qui vérifie l'identité :

$$Q(x) R(x) = P(x)$$

C'est donc un élément de $\overline{W}_{m,n}$.

Il reste à vérifier que :

$$\| \mathcal{J} - R \| = \overline{\rho}$$

Si $x \in Y$, on peut écrire :

$$\begin{aligned} |\mathcal{J}^{(i)}(x) - R^{(i)}(x)| &\leq |\mathcal{J}^{(i)}(x) - R_k^{(i)}(x)| + |R^{(i)}(x) - R_k^{(i)}(x)| \\ &\leq \max_{x \in I} |\mathcal{J}^{(i)}(x) - R_k^{(i)}(x)| + |R^{(i)}(x) - R_k^{(i)}(x)| \end{aligned}$$

Donc, quels que soient les points $x_0, x_1, \dots, x_p$ dans Y , on a :

$$\sum_{i=0}^p K_i |\mathcal{J}^{(i)}(x_i) - R^{(i)}(x_i)| \leq \| \mathcal{J} - R_k \| + \sum_{i=0}^p K_i |R^{(i)}(x_i) - R_k^{(i)}(x_i)|$$

et, en passant à la limite pour $k \rightarrow \infty$:

$$(20) \quad \sum_{i=0}^p K_i |\mathcal{J}^{(i)}(x_i) - R^{(i)}(x_i)| \leq \overline{\rho}$$

quels que soient $x_0, x_1, \dots, x_p$ dans Y .

Considérons maintenant $x \in I - Y$ et prenons une suite de points $\xi_j \in Y$ tels que :

$$|\mathcal{J}^{(i)}(x) - \mathcal{J}^{(i)}(\xi_j)| \leq \varepsilon_j \quad i = 0, 1, \dots, p$$

$$|R^{(i)}(x) - R^{(i)}(\xi_j)| \leq \varepsilon_j \quad i = 0, 1, \dots, p$$

avec : $\lim_{j \rightarrow \infty} \varepsilon_j = 0$

On peut alors écrire :

$$\begin{aligned} |\mathcal{J}^{(i)}(x) - R^{(i)}(x)| &\leq |\mathcal{J}^{(i)}(x) - \mathcal{J}^{(i)}(\xi_j)| + |\mathcal{J}^{(i)}(\xi_j) - R^{(i)}(\xi_j)| \\ &\quad + |R^{(i)}(\xi_j) - R^{(i)}(x)| \end{aligned}$$

c'est-à-dire ;

$$\begin{aligned} |\mathcal{J}^{(i)}(x) - R^{(i)}(x)| &\leq |\mathcal{J}^{(i)}(\xi_j) - R^{(i)}(\xi_j)| + 2\varepsilon_j \\ &\leq \max_{x \in Y} |\mathcal{J}^{(i)}(x) - R^{(i)}(x)| + 2\varepsilon_j \end{aligned}$$

D'où, en passant à la limite pour $j \rightarrow \infty$:

$$(21) \quad |\mathcal{J}^{(i)}(x) - R^{(i)}(x)| \leq \max_{y \in Y} |\mathcal{J}^{(i)}(y) - R^{(i)}(y)| \quad \forall x \in I - Y$$

Donc, d'après les formules (20) et (21), quels que soient $x_0, x_1, \dots, x_p$ dans I , on a :

$$\sum_{i=0}^p K_i |\mathcal{J}^{(i)}(x_i) - R^{(i)}(x_i)| \leq \bar{\rho}$$

c'est-à-dire ;

$$\|\mathcal{J} - R\| \leq \bar{\rho}$$

Comme par ailleurs, on a évidemment $\|\mathcal{J} - R\| \geq \bar{\rho}$, cela prouve que :

$$\|\mathcal{J} - R\| = \bar{\rho}$$

ce qui termine la démonstration d'existence.

On a ensuite une extension du théorème de La Vallée-Poussin :

Soit $R \in W_{m,n}$. S'il existe $(p+1)$ ensembles $D_0, D_1, \dots, D_p$ de points

de I tels que :

$$a) \quad f^{(i)}(x) - R^{(i)}(x) \neq 0 \quad \forall x \in D_i \quad (i = 0, 1, \dots, p)$$

$$b) \quad \sum_{i=0}^p K_i \inf_{x \in D_i} \{ [f^{(i)}(x) - R^{(i)}(x)] \operatorname{sign} [f^{(i)}(x) - R^{(i)}(x)] \} \leq 0$$

quel que soit $T \in W_{m,n}$

alors :

$$\rho \geq \sum_{i=0}^p K_i \inf_{x \in D_i} |f^{(i)}(x) - R^{(i)}(x)|$$

Si cette inégalité n'était pas vérifiée, il existerait un élément $R \in W_{m,n}$ tel que :

$$\|f - R_1\| < \sum_{i=0}^p K_i \inf_{x \in D_i} |f^{(i)}(x) - R^{(i)}(x)|$$

On aurait donc :

$$(22) \quad \sum_{i=0}^p K_i |f^{(i)}(x_i) - R^{(i)}(x_i)| < \sum_{i=0}^p K_i |f^{(i)}(x_i) - R^{(i)}(x_i)|$$

quels que soient $x_0 \in D_0, x_1 \in D_1, \dots, x_p \in D_p$.

Posons alors :

$$\begin{aligned} A &= \sum_{i=0}^p K_i [R_1^{(i)}(x_i) - R^{(i)}(x_i)] \operatorname{sign} [f^{(i)}(x_i) - R^{(i)}(x_i)] \\ A &= \sum_{i=0}^p K_i |f^{(i)}(x_i) - R^{(i)}(x_i)| - \sum_{i=0}^p K_i \varepsilon_i |f^{(i)}(x_i) - R_1^{(i)}(x_i)| \end{aligned}$$

avec $\varepsilon = \pm 1$. On en déduit :

$$A \geq \sum_{i=0}^p K_i |f^{(i)}(x_i) - R^{(i)}(x_i)| - \sum_{i=0}^p K_i |f^{(i)}(x_i) - R_1^{(i)}(x_i)|$$

La formule (22) montre que $A > 0$ quels que soient $x_0 \in D_0, x_1 \in D_1, \dots, x_p \in D_p$, ce qui est en contradiction avec l'hypothèse b). On a donc bien :

$$\rho \geq \sum_{i=0}^p K_i \inf_{x \in D_i} |f^{(i)}(x) - R^{(i)}(x)|$$

Pour $R \in W_{m,n}$, notons E_i ($i=0,1,\dots,p$) l'ensemble suivant :

$$E_i = \{x \in I \mid f^{(i)}(x) - R^{(i)}(x) = \pm \max_{t \in I} |f^{(i)}(t) - R^{(i)}(t)|\}$$

Du théorème de La Vallée-Poussin, on déduit une condition suffisante pour que $R \in W_{m,n}$ soit optimal pour f :

Pour que $R \in W_{m,n}$ soit optimal pour f , il suffit que l'on ait :

$$\sum_{i=0}^p K_i \inf_{x \in E_i} \{ [T^{(i)}(x) - R^{(i)}(x)] \operatorname{sign} [f^{(i)}(x) - R^{(i)}(x)] \} \leq 0$$

quel que soit $T \in W_{m,n}$.

Dans le cas où $p=0$ et où Ω est convexe, on avait montré (théorème 4) que cette condition est nécessaire. Dans le cas présent, il n'est pas possible d'utiliser une démonstration inspirée de celle du théorème 4. En effet, on avait posé :

$$R(x) = \frac{P(x)}{Q(x)}, \quad R_0(x) = \frac{P_0(x)}{Q_0(x)}$$

$$R_1(x) = \frac{(1-\lambda)P(x) + \lambda P_0(x)}{(1-\lambda)Q(x) + \lambda Q_0(x)}$$

On en déduisait :

$$R_1(x) - R(x) = \frac{\lambda Q_0(x)[R_0(x) - R(x)]}{(1-\lambda)Q(x) + \lambda Q_0(x)}$$

et on utilisait le fait que $R_1(x) - R(x)$ et $R_0(x) - R(x)$ ont même signe, propriété qui n'est pas en général vérifiée par leurs dérivées successives.

Mais par contre on a encore l'analogue du corollaire du théorème 4 :

Pour que $R \in W_{m,n}$ soit optimal pour f , il suffit qu'il existe $r \geq 1$ points $(x_{0,j}, x_{1,j}, \dots, x_{p,j})$ de $E_0 \times E_1 \times \dots \times E_p$ et r nombres $\lambda_j > 0$ tels que :

$$\sum_{j=1}^r \lambda_j \left\{ \sum_{i=0}^p K_i [T^{(i)}(x_{i,j}) - R^{(i)}(x_{i,j})] \operatorname{sign} [f^{(i)}(x_{i,j}) - R^{(i)}(x_{i,j})] \right\} \leq 0$$

quel que soit $T \in W_{m,n}$.

En effet, cette condition montre qu'il n'existe aucun élément $T \in W_{m,n}$ tel que :

$$\sum_{i=0}^p K_i \{ [T^{(i)}(x_i) - R^{(i)}(x_i)] \operatorname{sign} [J^{(i)}(x_i) - R^{(i)}(x_i)] \} > 0$$

quel que soit $(x_0, x_1, \dots, x_p) \in E_0 \times E_1 \times \dots \times E_p$,

ce qui prouve que R est optimal pour J .

Chapitre II

APPROXIMATION RATIONNELLE AVEC CONTRAINTES DU TYPE EGALITES ET INEGALITES LINEAIRES

Dans ce chapitre, on étudie l'approximation d'une fonction continue $f(x)$ par des "fractions rationnelles généralisées" :

$$R(x) = \frac{\sum_{j=0}^m a_j g_j(x)}{\sum_{j=0}^n b_j h_j(x)}$$

dont les coefficients vérifient des égalités et inégalités linéaires :

$$\sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i \quad i = 1, \dots, K$$

$$\sum_{j=0}^m \alpha_j \alpha_j(t) + \sum_{j=0}^n b_j \beta_j(t) \geq c(t) \quad \forall t \in G$$

Le problème est posé dans le paragraphe 1. Au paragraphe 2, on examine ce que deviennent dans ce cas particulier les résultats du chapitre I concernant l'existence, la convexité de l'ensemble des éléments optimaux et la généralisation du théorème de la Vallée-Poussin. La caractérisation des éléments optimaux est étudiée au paragraphe 3. Au paragraphe 4, on donne une condition nécessaire et suffisante pour qu'il y ait unicité de la solution et dans le paragraphe 5, on aborde les propriétés d'alternance.

1 - Introduction.

Comme au chapitre précédent, on considère un espace compact X et $C[X]$ espace vectoriel des fonctions continues sur X , à valeurs réelles. On se donne dans $C[X]$ des fonctions $g_0(x), \dots, g_m(x)$, $h_0(x), \dots, h_n(x)$ et on suppose les $g_i(x)$ d'une part et les $h_i(x)$ d'autre part linéairement indépendantes sur X .

On désigne par $W_{m,n}$ la classe des fonctions :

$$R(x) = \frac{\sum_{j=0}^m a_j g_j(x)}{\sum_{j=0}^n b_j h_j(x)}$$

où les a_j, b_j sont des nombres réels vérifiant les conditions :

$$(1) \quad \sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i \quad i = 1, \dots, K \quad (K \geq 1)$$

$$(2) \quad \sum_{j=0}^m a_j \alpha_j(t) + \sum_{j=0}^n b_j \beta_j(t) \geq c(t) \quad \forall t \in G$$

$$(3) \quad \sum_{j=0}^n b_j h_j(x) > 0 \quad \text{sur } X$$

Les nombres $\alpha_{ij}, \beta_{ij}, c_i$ sont donnés et on suppose que les vecteurs $V_i (i = 1, \dots, K)$ de composantes $(\alpha_{i0}, \alpha_{i1}, \dots, \alpha_{im}, \beta_{i0}, \beta_{i1}, \dots, \beta_{in})$ sont linéairement indépendants.

G désigne un compact arbitraire et les fonctions $\alpha_j(t), \beta_j(t)$ et $c(t)$ sont données dans $C[G]$.

Enfin, on suppose que la fonction $c(t)$ vérifie la condition de régularité suivante :

Il existe un système de $(m + n + 2)$ nombres réels $\bar{a}_0, \bar{a}_1, \dots, \bar{a}_m, \bar{b}_0, \bar{b}_1, \dots, \bar{b}_n$ tels que :

$$(4) \quad \sum_{j=0}^m \alpha_{ij} \bar{a}_j + \sum_{j=0}^n \beta_{ij} \bar{b}_j = c_i \quad i = 1, \dots, K$$

$$\sum_{j=0}^m \bar{a}_j \alpha_j(t) + \sum_{j=0}^n \bar{b}_j \beta_j(t) > c(t) \quad \forall t \in G$$

Problème: Il s'énonce comme au chapitre précédent :

étant donné $f \in C[X]$, $f \notin W_{m,n}$, trouver dans $W_{m,n}$ un élément R^* tel que :

$$\|f - R^*\| \leq \|f - R\| \quad \forall R \in W_{m,n}$$

Remarque : L'ensemble Ω défini au chapitre précédent est ici l'ensemble des points $(a_0, \dots, a_m, b_0, \dots, b_n)$ de R^{m+n+2} qui vérifient les relations (1) et (2). C'est donc un ensemble convexe fermé. La condition de régularité signifie que Ω n'est pas réduit à une variété linéaire (c'est-à-dire que, compte tenu des égalités (1), les inégalités (2) ne se réduisent pas à des égalités). Ω est alors contenu dans une variété linéaire de dimension minimum uniquement déterminée et l'intérieur de Ω relativement à cette variété est non vide (EGGLESTON [1958], Théorème 4). Donc il existe un point $(\bar{a}_0, \bar{a}_1, \dots, \bar{a}_m, \bar{b}_0, \bar{b}_1, \dots, \bar{b}_n)$ vérifiant les relations (4).

Notations :

Comme au chapitre précédent, nous utiliserons, pour $R \in W_{m,n}$, les notations suivantes :

$$y_R(x) = f(x) - R(x)$$

$$\rho = \inf_{R \in W_{m,n}} \|f - R\|$$

E_R désigne encore l'ensemble des (e) - points de $f - R$.

$\mathcal{D}_R$ désigne la classe des fonctions :

$$\Phi(x) = \sum_{j=0}^m a'_j g_j(x) - R(x) \sum_{j=0}^n b'_j h_j(x)$$

dont les coefficients a'_j, b'_j sont des nombres réels vérifiant (1) et (2).

Enfin, Δ_R désigne la classe des fonctions :

$$\Phi(x) = \sum_{j=0}^m a'_j g_j(x) - R(x) \sum_{j=0}^n b'_j h_j(x)$$

dont les coefficients a'_j, b'_j vérifient seulement la condition (1).

2 - Existence d'un élément optimal.

Comme au chapitre précédent, désignons par $\bar{W}_{m,n}$ la classe des fonctions $R \in C[X]$ qui vérifient une identité de la forme :

$$R(x) \sum_{i=0}^n b_i h_i(x) \equiv \sum_{i=0}^m a_i g_i(x)$$

où les a_i, b_i sont des nombres réels non tous nuls vérifiant les conditions (1) et (2).

On a tout d'abord le théorème d'existence suivant :

Théorème 1 : si $c_i = 0$ ($i = 1, \dots, K$), $c(t) \leq 0$ ($\forall t \in G$) et si les fonctions $h_0(x), h_1(x), \dots, h_n(x)$ sont linéairement indépendantes sur tout sous-ensemble ouvert de X , alors toute fonction $f \in C[X]$ possède dans $\bar{W}_{m,n}$ au moins un élément optimal.

Ce résultat se déduit immédiatement du théorème I-1 puisque, lorsque $c_i = 0$ et $c(t) \leq 0$, l'hypothèse (H2) de ce théorème est vérifiée.

L'ensemble Ω étant convexe, le théorème I-2 s'applique. D'autre part, le théorème I-3 (généralisation du théorème de la Vallée-Poussin) ne comportant aucune hypothèse sur Ω est bien entendu valable ici. Ce qui nous donne les deux résultats suivants :

Théorème 2 : l'ensemble des points de R^{m+n+2} dont les coordonnées $(a_0, a_1, \dots, a_m, b_0, b_1, \dots, b_n)$ sont les coefficients des éléments $R \in W_{m,n}$ optimaux pour f est convexe.

Théorème 3 : soit $R \in W_{m,n}$. S'il existe un ensemble $D \subset X$ tel que :

$$a) \quad f(x) - R(x) \neq 0 \quad \forall x \in D,$$

$$b) \quad \inf_{\Phi \in \mathcal{D}} \Phi(x) [f(x) - R(x)] \leq 0 \quad \forall \Phi \in \mathcal{D}_R,$$

alors, on a :

$$\rho \geq \inf_{x \in D} |f(x) - R(x)|$$

3 - Caractérisation des éléments optimaux.

L'ensemble Ω étant convexe, le théorème I-4 va tout d'abord nous fournir une condition nécessaire et suffisante pour que $R \in W_{m,n}$ soit optimal :

Théorème 4 : pour que $R \in W_{m,n}$ soit optimal pour f , il faut et il suffit que l'on ait :

$$\inf_{x \in E_R} \Phi(x) [f(x) - R(x)] \leq 0 \quad \text{pour tout} \quad \Phi \in \mathcal{D}_R$$

Compte tenu de la définition explicite de $\mathfrak{Q}$ par les relations (1) et (2), on peut donner une autre forme de ce théorème de caractérisation :

Théorème 5 : pour que $R \in W_{m,n}$ soit optimal pour f , il faut et il suffit qu'il existe des (e)-points $x_1, \dots, x_r$ de $f - R$, des points $t_1, \dots, t_s$ de G et des nombres $\lambda_1, \dots, \lambda_r, \mu_1, \dots, \mu_s, \gamma_1, \dots, \gamma_K$ tels que :

$$(5) \quad r \geq 1, \quad r + s \leq 1 + \dim(\Delta_R)$$

$$(6) \quad \mu_i > 0 \quad (i = 1, \dots, s), \quad \lambda_i [f(x_i) - R(x_i)] > 0 \quad (i = 1, \dots, r)$$

$$(7) \quad \sum_{i=1}^r \lambda_i \begin{bmatrix} g_0(x_i) \\ \vdots \\ g_m(x_i) \\ -R(x_i) \cdot h_0(x_i) \\ \vdots \\ -R(x_i) \cdot h_n(x_i) \\ 0 \end{bmatrix} + \sum_{i=1}^s \mu_i \begin{bmatrix} \alpha_0(t_i) \\ \vdots \\ \alpha_m(t_i) \\ \beta_0(t_i) \\ \vdots \\ \beta_n(t_i) \\ c(t_i) \end{bmatrix} + \sum_{i=1}^K \gamma_i \begin{bmatrix} \alpha_{i0} \\ \vdots \\ \alpha_{im} \\ \beta_{i0} \\ \vdots \\ \beta_{in} \\ c_i \end{bmatrix} = 0$$

1°) Supposons que $R \in W_{m,n}$ vérifie les conditions (5), (6), (7) ci dessus. Considérons dans $\mathcal{D}_R$ un élément quelconque :

$$\Phi(x) = \sum_{j=0}^m a_j g_j(x) - R(x) \sum_{j=0}^n b_j h_j(x)$$

Multiplions la relation matricielle (7) par la matrice ligne $(a_0, a_1, \dots, a_m, b_0, b_1, \dots, b_n, -1)$. On obtient :

$$(8) \quad \sum_{i=1}^r \lambda_i \Phi(x_i) + \sum_{i=1}^s \mu_i \left[\sum_{j=0}^m a_j \alpha_j(t_i) + \sum_{j=0}^n b_j \beta_j(t_i) - c(t_i) \right] + \sum_{i=1}^K \gamma_i \left[\sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j - c_i \right] = 0$$

Mais puisque Φ est un élément de $\mathcal{D}_R$, on a :

$$(9) \quad \begin{cases} \sum_{j=0}^m a_j \alpha_j(t_i) + \sum_{j=0}^n b_j \beta_j(t_i) \geq c(t_i) & i = 1, \dots, s \\ \sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i & i = 1, \dots, K \end{cases}$$

De (8) et (9), on déduit :

$$\sum_{i=1}^r \lambda_i \Phi(x_i) \leq 0 \quad \forall \Phi \in \mathcal{D}_R$$

Compte tenu de (6), ceci prouve, d'après le corollaire du théorème I-4 que R est optimal.

2°) Réciproquement, soit $R(x) = \sum a_j^* g_j(x) / \sum b_j^* h_j(x)$ optimal pour f dans $W_{m,n}$. D'après le théorème 4, le système suivant d'égalités et inégalités linéaires par rapport aux inconnues a_j, b_j est incompatible :

$$(10) \quad \left\{ \begin{array}{ll} \forall x \in E_R & \left[\sum_{j=0}^m a_j g_j(x) - R(x) \sum_{j=0}^n b_j h_j(x) \right] > 0 \\ \forall t \in G & \sum_{j=0}^m a_j \alpha_j(t) + \sum_{j=0}^n b_j \beta_j(t) \geq c(t) \\ i = 1, \dots, K & \sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i \end{array} \right.$$

Soit $\Phi(x)$ un élément de Δ_R :

$$(11) \quad \Phi(x) = \sum_{j=0}^m a_j g_j(x) - R(x) \sum_{j=0}^n b_j h_j(x)$$

A cet élément Φ , associons la fonction $\phi(t)$ définie par :

$$(12) \quad \phi(t) = \sum_{j=0}^m a_j \alpha_j(t) + \sum_{j=0}^n b_j \beta_j(t)$$

Posons d'autre part :

$$(13) \quad \left\{ \begin{array}{l} \Phi^*(x) = \sum_{j=0}^m a_j^* g_j(x) - R(x) \sum_{j=0}^n b_j^* h_j(x) \equiv 0 \\ \phi^*(t) = \sum_{j=0}^m a_j^* \alpha_j(t) + \sum_{j=0}^n b_j^* \beta_j(t) \end{array} \right.$$

Ce qui permet d'écrire (11) et (12) sous la forme :

$$(14) \quad \left\{ \begin{array}{l} \Phi(x) = \sum_{j=0}^m (a_j - a_j^*) g_j(x) - R(x) \sum_{j=0}^n (b_j - b_j^*) h_j(x) \\ \phi(t) = \sum_{j=0}^m (a_j - a_j^*) \alpha_j(t) + \sum_{j=0}^n (b_j - b_j^*) \beta_j(t) + \phi^*(t) \end{array} \right.$$

Soit $\Phi_1(x), \dots, \Phi_M(x)$ une base de l'espace Δ_R et soient $\varphi_1(x), \dots, \varphi_M(t)$ les fonctions associées aux $\Phi_i(x)$ à l'aide des formules (11) et (12).

Les formules (14) s'écrivent alors :

$$(15) \quad \begin{cases} \Phi(x) = \sum_1^M \tau_i \Phi_i(x) \\ \varphi(t) = \varphi^*(t) + \sum_1^M \tau_i \varphi_i(t) \end{cases}$$

D'après (10), le système suivant d'inégalités linéaires par rapport aux inconnues τ_i est incompatible :

$$(16) \quad \begin{cases} y_R(x) \sum_1^M \tau_i \Phi_i(x) > 0 & \forall x \in E_R \\ \varphi^*(t) - \alpha(t) + \sum_1^M \tau_i \varphi_i(t) \geq 0 & \forall t \in G \end{cases}$$

Il en est évidemment de même pour le système :

$$(17) \quad \begin{cases} y_R(x) [\tau_1 \Phi_1(x) + \dots + \tau_M \Phi_M(x)] > 0 & \forall x \in E_R \\ \tau_0 [\varphi^*(t) - \alpha(t)] + [\tau_1 \varphi_1(t) + \dots + \tau_M \varphi_M(t)] > 0 & \forall t \in G \\ \tau_0 > 0 \end{cases}$$

Donc d'après le théorème 0-2, l'origine de R^{M+1} appartient à l'enveloppe convexe de l'ensemble suivant :

$$\left\{ \begin{bmatrix} 0 \\ y_R(x) \Phi_1(x) \\ \vdots \\ y_R(x) \Phi_M(x) \end{bmatrix}, x \in E_R \right\} \cup \left\{ \begin{bmatrix} \varphi^*(t) - \alpha(t) \\ \varphi_1(t) \\ \vdots \\ \varphi_M(t) \end{bmatrix}, t \in G \right\} \cup \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

D'après le théorème de CARATHÉODORY (théorème 0-1), il existe des (e)-points $x_1, \dots, x_r$, et des points $t_1, \dots, t_s$, de G tels que :

$$(18) \quad \sum_{i=1}^{r'} \lambda_i \begin{bmatrix} 0 \\ \Phi_1(x_i) \\ \vdots \\ \Phi_M(x_i) \end{bmatrix} + \sum_{i=1}^{s'} \mu_i \begin{bmatrix} \varphi^*(t_i) - c(t_i) \\ \varphi_1(t_i) \\ \vdots \\ \varphi_M(t_i) \end{bmatrix} + \nu \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$$

avec :

$$\begin{cases} r' \geq 0, s' \geq 0, r' + s' \leq M + 2 \\ \lambda_i y_R(x_i) > 0, \mu_i > 0, \nu \geq 0 \end{cases}$$

La première ligne de l'égalité matricielle (18) s'écrit :

$$\sum_{i=1}^{s'} \mu_i [\varphi^*(t_i) - c(t_i)] + \nu = 0$$

Comme $R \in W_{m,n}$, on a : $\varphi^*(t) \geq c(t) \quad \forall t \in G$
et donc :

$$(19) \quad \begin{aligned} \varphi^*(t_i) &= c(t_i) & i &= 1, \dots, s' \\ \nu &= 0 \end{aligned}$$

La relation (18) peut alors s'écrire :

$$\sum_{i=1}^{r'} \lambda_i \Phi_j(x_i) + \sum_{i=1}^{s'} \mu_i \varphi_j(t_i) = 0 \quad j = 1, \dots, M$$

ce qui prouve que le système d'inéquations suivant est incompatible :

$$(20) \quad \begin{cases} y_R(x_i) [\tau_1 \Phi_1(x_i) + \dots + \tau_M \Phi_M(x_i)] > 0 & i = 1, \dots, r' \\ \tau_1 \varphi_1(t_i) + \dots + \tau_M \varphi_M(t_i) > 0 & i = 1, \dots, s' \end{cases}$$

En raisonnant sur le système (20) comme sur le système (17), on en déduit qu'il existe des (e)-points $x_1, \dots, x_r$ et des points $t_1, \dots, t_s$ de G tels que :

$$(21) \quad \sum_{i=1}^r \lambda_i \begin{bmatrix} \Phi_1(x_i) \\ \vdots \\ \Phi_M(x_i) \end{bmatrix} + \sum_{i=1}^s \mu_i \begin{bmatrix} \varphi_1(t_i) \\ \vdots \\ \varphi_M(t_i) \end{bmatrix} = 0$$

avec :

$$\begin{cases} r \geq 0, s \geq 0, 1 \leq r + s \leq M + 1 \\ \lambda_i y_R(x_i) > 0, \mu_i > 0 \end{cases}$$

D'où l'on déduit :

$$\sum_{i=1}^r \lambda_i \left[\sum_{j=1}^M \tau_j \Phi_j(x_i) \right] + \sum_{i=1}^s \mu_i \left[\sum_{j=1}^M \tau_j \varphi_j(t_i) \right] = 0$$

Ce qui, en tenant compte de (15) et (19) s'écrit :

$$(22) \quad \sum_{i=1}^r \lambda_i \Phi(x_i) + \sum_{i=1}^s \mu_i [\varphi(t_i) - c(t_i)] = 0 \quad \forall \Phi \in \Delta_R$$

Vérifions maintenant que $r \geq 1$. Si on avait $r = 0$, la relation (22) s'écrit simplement :

$$\sum_{i=1}^s \mu_i [\varphi(t_i) - c(t_i)] = 0 \quad \forall \Phi \in \Delta_R$$

Or d'après la condition de régularité, il existe dans Δ_R un élément $\Phi(x)$ tel que :

$$\Phi(t) > c(t) \quad \forall t \in G$$

Comme $\mu_i > 0$ et $s \geq 1$, il y a une contradiction et l'on a donc bien : $r \geq 1$.

Nous venons de vérifier les relations (5) et (6). Il nous reste à transformer la relation (22) pour obtenir (7). Or (22) signifie que toute solution $(a_0, \dots, a_m, b_0, \dots, b_n)$ du système :

$$(23) \quad \sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i \quad i = 1, \dots, K$$

vérifie l'équation supplémentaire :

$$(24) \quad \sum_{i=1}^r \lambda_i \left[\sum_{j=0}^m \alpha_j g_j(x_i) - R(x_i) \sum_{j=0}^n b_j h_j(x_i) \right] + \sum_{i=1}^s \mu_i \left[\sum_{j=0}^m \alpha_j \alpha_j(t_i) + \sum_{j=0}^n b_j \beta_j(t_i) - c(t_i) \right] = 0$$

que nous écrirons plus simplement :

$$(25) \quad \sum_{j=0}^m A_j a_j + \sum_{j=0}^n B_j b_j = D$$

$$(29) \quad \begin{aligned} D + \sum_{l=1}^K \gamma_l c_l &= 0 \\ A_j + \sum_{l=1}^K \gamma_l \alpha_{lj} &= 0 \end{aligned} \quad j = q, \dots, m$$

$$(30) \quad B_j + \sum_{l=1}^K \gamma_l \beta_{lj} = 0 \quad j = K - q, \dots, n$$

D'après (27), la relation (29) est vraie pour j variant de 0 à m et la relation (30) pour j variant de 0 à n . En tenant compte des définitions de A_j, B_j, D , on obtient finalement :

$$\left\{ \begin{aligned} \sum_{l=1}^r \lambda_l g_j(x_l) + \sum_{l=1}^s \mu_l \alpha_j(t_l) + \sum_{l=1}^K \gamma_l \alpha_{lj} &= 0 & j = 0, \dots, m \\ \sum_{l=1}^r -\lambda_l R(x_l) h_j(x_l) + \sum_{l=1}^s \mu_l \beta_j(t_l) + \sum_{l=1}^K \gamma_l \beta_{lj} &= 0 & j = 0, \dots, n \\ \sum_{l=1}^s \mu_l c(t_l) + \sum_{l=1}^K \gamma_l c_l &= 0 \end{aligned} \right.$$

On a bien obtenu la formule (7).

Remarque 1 : l'hypothèse $r + s \leq 1 + \dim(\Delta_R)$ n'intervient pas pour démontrer que la condition est suffisante. Par suite, le théorème 5 reste valable lorsqu'on remplace $1 + \dim(\Delta_R)$ par un nombre plus grand. On peut prendre :

$$r + s \leq m + n + 3 - K \quad \text{dans le cas général}$$

$$r + s \leq m + n + 2 - K \quad \text{si } c_l = 0 \quad (l = 1, \dots, K)$$

En effet, tout élément de Δ_R s'écrit :

$$\Phi(x) = \sum_0^m a_j g_j(x) - R(x) \sum_0^n b_j h_j(x)$$

avec :

$$(31) \quad \sum_{j=0}^m \alpha_{lj} a_j + \sum_{j=0}^n \beta_{lj} b_j = c_l \quad (l = 1, \dots, K)$$

Ce système d'équations permet de calculer K des coefficients a_j, b_j en fonction des autres :

$$\begin{cases} a_i = \varepsilon_i + \sum_{j=q}^m \gamma_{ij} a_j + \sum_{j=K-q}^n \delta_{ij} b_j & i = 0, \dots, q-1 \\ b_i = \varepsilon'_i + \sum_{j=q}^m \gamma'_{ij} a_j + \sum_{j=K-q}^n \delta'_{ij} b_j & i = 0, \dots, K-q-1 \end{cases}$$

Ce qui permet d'écrire tout élément de Δ_R sous la forme :

$$\Phi(x) = \Phi_0(x) + \sum_{j=q}^m a_j \Phi_{j-q+1}(x) + \sum_{j=K-q}^n b_j \Phi_{m+2+j-K}(x)$$

Mais on a :

$$(32) \quad \Phi^*(x) = \Phi_0(x) + \sum_{j=q}^m a_j^* \Phi_{j-q+1}(x) + \sum_{j=K-q}^n b_j^* \Phi_{m+2+j-K}(x) = 0$$

D'où :

$$\Phi(x) = \sum_{j=q}^m (a_j - a_j^*) \Phi_{j-q+1}(x) + \sum_{j=K-q}^n (b_j - b_j^*) \Phi_{m+2+j-K}(x)$$

Ce qui prouve que dans le cas général, on a :

$$\dim(\Delta_R) \leq m + n + 2 - K$$

Dans le cas où $c_i = 0$ ($i = 1, \dots, K$), on a :

$$\varepsilon_i = \varepsilon'_i = 0 \quad \Phi_0(x) = 0$$

et (32) montre que les $\Phi_i(x)$ sont linéairement dépendants et par suite :

$$\dim(\Delta_R) \leq m + n + 1 - K$$

Remarque 2 : D'après la formule (19), on a :

$$(33) \quad \sum_{j=0}^m a_j^* \alpha_j(t_i) + \sum_{j=0}^n b_j^* \beta_j(t_i) = c(t_i) \quad (i = 1, \dots, s)$$

Si $s = 0$ (c'est-à-dire s'il n'y a pas de points t_i), cela signifie que $R(x)$ est optimal pour $f(x)$ dans la classe $W'_{m,n}$ des fonctions qui vérifient seulement les contraintes :

$$(34) \quad \sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i \quad (i = 1, \dots, K)$$

Si $s \geq 1$, les relations (33) montrent que le point de coordonnées $(a_0^*, \dots, a_m^*, b_0^*, \dots, b_n^*)$ appartient à la frontière de Ω (relativement à la variété linéaire V' définie par (34)).

On retrouve ainsi le résultat du corollaire du théorème I - 5.

Remarque 3 : Le théorème 5 peut encore s'énoncer sous la forme suivante :

Pour que $R(x) = \sum a_j^* g_j(x) / \sum b_j^* h_j(x)$ soit optimal dans $W_{m,n}$,

il faut et il suffit qu'il existe des (e) -points $x_1, \dots, x_r$ ($r \geq 1$) de $\mathcal{F} - R$, des points $t_1, \dots, t_s$ de G et des nombres $\lambda_i, \mu_i, \gamma_i$ tels que :

$$r + s \leq 1 + \dim(\Delta_R)$$

$$\mu_i > 0, \quad \lambda_i [\mathcal{F}(x_i) - R(x_i)] > 0$$

$$\sum_{j=0}^m a_j^* \alpha_j(t_i) + \sum_{j=0}^n b_j^* \beta_j(t_i) = c(t_i) \quad i = 1, \dots, s$$

$$\sum_{i=1}^r \lambda_i \begin{bmatrix} g_0(x_i) \\ \vdots \\ g_m(x_i) \\ -R(x_i)h_0(x_i) \\ \vdots \\ -R(x_i)h_n(x_i) \end{bmatrix} + \sum_{i=1}^s \mu_i \begin{bmatrix} \alpha_0(t_i) \\ \vdots \\ \alpha_m(t_i) \\ \beta_0(t_i) \\ \vdots \\ \beta_n(t_i) \end{bmatrix} + \sum_{i=1}^K \gamma_i \begin{bmatrix} \alpha_{i0} \\ \vdots \\ \alpha_{im} \\ \beta_{i0} \\ \vdots \\ \beta_{in} \end{bmatrix} = 0$$

Pour $n = 0$, on retrouve le résultat donné dans le cas polynomial par GOLSHTEIN [1961].

Remarque 4 : On peut prendre les conditions d'inégalité sous une forme légèrement différente et imposer aux coefficients des éléments de $W_{m,n}$ les conditions suivantes :

$$\left\{ \begin{array}{l} \sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i \quad i = 1, \dots, K \\ \left| \sum_{j=0}^m a_j \alpha_j(t) + \sum_{j=0}^n b_j \beta_j(t) - \Gamma(t) \right| \leq M \quad \forall t \in G \end{array} \right.$$

Cette condition d'inégalité s'écrit encore :

$$\begin{cases} \sum_{j=0}^m a_j \alpha_j(t) + \sum_{j=0}^n b_j \beta_j(t) \geq \Gamma(t) - M & \forall t \in G \\ \sum_{j=0}^m a_j [-\alpha_j(t)] + \sum_{j=0}^n b_j [-\beta_j(t)] \geq -\Gamma(t) - M & \forall t \in G \end{cases}$$

En appliquant le théorème 5, sous la forme donnée dans la remarque 3 ci-dessus, on obtient le résultat suivant :

Pour que $R(x) = \sum a_j^* g_j(x) / \sum b_j^* h_j(x)$ soit optimal pour f dans $W_{m,n}$, il faut et il suffit qu'il existe des (e)-points $x_1, \dots, x_r$ ($r \geq 1$) de $f - R$, des points $t_1, \dots, t_s, t'_1, \dots, t'_{s'}$ de G et des nombres $\lambda_i, \mu_i, \mu'_i, \gamma_i$ tels que :

$$r + s + s' \leq 1 + \dim(\Delta_R), \quad s \geq 0, \quad s' \geq 0$$

$$\mu_i > 0, \quad \mu'_i < 0, \quad \lambda_i [f(x_i) - R(x_i)] > 0$$

$$\sum_0^m a_j^* \alpha_j(t_i) + \sum_0^n b_j^* \beta_j(t_i) - \Gamma(t_i) = -M \quad (i = 1, \dots, s)$$

$$\sum_0^m a_j^* \alpha_j(t'_i) + \sum_0^n b_j^* \beta_j(t'_i) - \Gamma(t'_i) = +M \quad (i = 1, \dots, s')$$

$$\sum_{i=1}^r \lambda_i \begin{bmatrix} g_0(x_i) \\ \vdots \\ g_m(x_i) \\ R(x_i)h_0(x_i) \\ \vdots \\ R(x_i)h_n(x_i) \end{bmatrix} + \sum_{i=1}^s \mu_i \begin{bmatrix} \alpha_0(t_i) \\ \vdots \\ \alpha_m(t_i) \\ \beta_0(t_i) \\ \vdots \\ \beta_n(t_i) \end{bmatrix} + \sum_{i=1}^{s'} \mu'_i \begin{bmatrix} \alpha_0(t'_i) \\ \vdots \\ \alpha_m(t'_i) \\ \beta_0(t'_i) \\ \vdots \\ \beta_n(t'_i) \end{bmatrix} + \sum_{i=1}^K \gamma_i \begin{bmatrix} \alpha_{i0} \\ \vdots \\ \alpha_{im} \\ \beta_{i0} \\ \vdots \\ \beta_{in} \end{bmatrix} = 0$$

Lorsque $K = 0$ et $n = 0$, on retrouve le résultat donné dans le cas polynomial par TERRENOIRE ([1967]), proposition 7 - 2).

Du théorème 5 on déduit ensuite le résultat suivant :

Corollaire 1 : pour que $R \in W_{m,n}$ soit optimal pour f , il faut et il suffit qu'il existe des (e)-points $x_1, \dots, x_r$ ($r \geq 1$) de $f - R$, des points $t_1, \dots, t_s$ de G et des nombres λ_i, μ_i tels que :

$$r + s \leq 1 + \dim(\Delta_R)$$

$$\mu_i > 0, \quad \lambda_i [f(x_i) - R(x_i)] > 0$$

Pour tout élément $\Phi(x) = \sum_0^m a_j g_j(x) - R(x) \sum_0^n b_j h_j(x)$ de Δ_R , on a :

$$\sum_{i=1}^r \lambda_i \Phi(x_i) + \sum_{i=1}^s \mu_i \left[\sum_{j=0}^m a_j \alpha_j(t_i) + \sum_{j=0}^n b_j \beta_j(t_i) - c(t_i) \right] = 0$$

Cette condition est nécessaire d'après la formule (22).

Elle est suffisante d'après la corollaire du théorème I - 4 car de la condition ci-dessus, on déduit immédiatement :

$$\sum_{i=1}^r \lambda_i \Phi(x_i) \leq 0 \quad \forall \Phi \in \mathcal{D}_R.$$

D'après la remarque 3, le théorème 5 peut encore s'énoncer sous la forme suivante :

Pour que $R(x) = \sum a_j^* g_j(x) / \sum b_j^* h_j(x)$ soit optimal pour f , il faut et il suffit qu'il existe des points $x_1, \dots, x_r$ ($r \geq 1$) de X , des points $t_1, \dots, t_s$ de G et une fonctionnelle linéaire L tels que :

$$r + s \leq 1 + \dim(\Delta_R)$$

$$\left\{ \begin{array}{l} \sum_{j=0}^m a_j^* \alpha_j(t_i) + \sum_{j=0}^n b_j^* \beta_j(t_i) = c(t_i) \quad i = 1, \dots, s \\ f(x_i) - R(x_i) = \varepsilon_i \|f - R\| \quad (\varepsilon_i = \pm 1) \\ \text{avec : } \begin{bmatrix} L(g_0) \\ \vdots \\ L(-R h_n) \end{bmatrix} + \sum_{i=1}^s \mu_i \begin{bmatrix} \alpha_0(t_i) \\ \vdots \\ \beta_n(t_i) \end{bmatrix} + \sum_{i=1}^K \gamma_i \begin{bmatrix} \alpha_{i0} \\ \vdots \\ \beta_{in} \end{bmatrix} = 0 \\ \text{et : } L(\Phi) = \sum_{i=1}^r \lambda_i \varepsilon_i \Phi(x_i) \quad (\lambda_i > 0, \sum \lambda_i = 1) \\ \mu_i > 0 \end{array} \right.$$

Cette fonctionnelle L vérifie :

$$L(f - R) = \|f - R\| \quad \text{et} \quad \|L\| = 1$$

Inversement, considérons une fonctionnelle linéaire :

$$L(\Phi) = \sum_1^r \lambda_i \Phi(x_i) \quad \text{avec} \quad \sum |\lambda_i| = 1$$

telle que :

$$L(f - R) = \|f - R\|$$

Cette condition montre que les x_i sont des (e) -points de $f - R$ tels que :

$$\lambda_i [f(x_i) - R(x_i)] > 0 \quad (i = 1, \dots, r)$$

Ce qui permet de mettre le théorème 5 sous la forme suivante :

Corollaire 2 : pour que $R(x) = \sum a_j^* g_j(x) / \sum b_j^* h_j(x)$ soit optimal pour f , il faut et il suffit qu'il existe une fonctionnelle linéaire L , des points $t_1, \dots, t_s$ de G et des nombres μ_i, γ_i tels que :

$$\sum_{j=0}^m a_j^* \alpha_j(t_i) + \sum_{j=0}^n b_j^* \beta_j(t_i) = o(t_i) \quad (i = 1, \dots, s)$$

$$L(\Phi) = \sum_{i=1}^r \lambda_i \Phi(x_i) \quad , \quad \sum |\lambda_i| = 1$$

$$L(f - R) = \|f - R\| \quad , \quad \|L\| = 1 \quad , \quad \mu_i > 0$$

$$\left\{ \begin{bmatrix} L(g_0) \\ \vdots \\ L(g_m) \\ L(-Rh_0) \\ \vdots \\ L(-Rh_n) \end{bmatrix} + \sum_{i=1}^s \mu_i \begin{bmatrix} \alpha_0(t_i) \\ \vdots \\ \alpha_m(t_i) \\ \beta_0(t_i) \\ \vdots \\ \beta_n(t_i) \end{bmatrix} + \sum_{i=1}^K \gamma_i \begin{bmatrix} \alpha_{i0} \\ \vdots \\ \alpha_{im} \\ \beta_{i0} \\ \vdots \\ \beta_{in} \end{bmatrix} \right\} = 0$$

Les $K + s$ derniers vecteurs ci-dessus sont indépendants.

On retrouve sous cette forme le résultat donné dans le cas polynomial (avec contraintes linéaires) par GOLSHEIN [1962].

Ce résultat peut aussi se mettre sous la forme :

Corollaire 3 : pour que $R \in W_{m,n}$ soit optimal, il faut et il suffit qu'il existe une fonctionnelle linéaire L telle que :

$$\begin{aligned} L(f - R) &= \|f - R\| \quad , \quad \|L\| = 1 \\ L(\Phi) &\leq 0 \quad \forall \Phi \in \mathcal{D}_R \quad (\text{voir p. 149}) \end{aligned}$$

Cette condition est nécessaire d'après ce qui précède et suffisante d'après le théorème I - 4.

4 - Unicité

Soit $\mathcal{B}$ l'ensemble des couples de fonctions (f, c) tels que :

- 1) $f(x) \in C(X)$, $c(t) \in C[G]$
- 2) $c(t)$ vérifie la condition de régularité.
- 3) $f(x)$ a au moins un élément optimal dans la classe $W_{m,n}$ (associée à la fonction $c(t)$ par les formules (1), (2), (3) du paragraphe 1).

Nous allons donner une condition nécessaire et suffisante pour que le problème de la meilleure approximation de $f(x)$ par les éléments de $W_{m,n}$ ait une solution unique pour tout couple $(f, c) \in \mathcal{B}$. Ce résultat généralise celui donné par GOLSHTEIN [1961, 1961 a]. Au cours de la démonstration, nous obtiendrons également une condition suffisante pour que l'élément $R \in W_{m,n}$ optimal pour f soit unique.

Théorème 6 : le problème posé a une solution unique pour tout élément $(f, c) \in \mathcal{B}$

si et seulement si toute fraction $R(x) = \sum a_j g_j(x) / \sum b_j h_j(x)$ telle que :

$$\left\{ \begin{aligned} &\sum_{j=0}^n b_j h_j(x) > 0 \quad \text{sur } X \\ &\sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i \quad (i = 1, \dots, K) \end{aligned} \right.$$

possède la propriété (P) énoncée ci-dessous

Propriété (P) : il n'existe pas de points $x_1, \dots, x_r$ de X , de points $t_1, \dots, t_s$ de G et de nombres $\lambda_i, \mu_i, \gamma_i$ tels que :

$$\mu_i > 0 \quad , \quad \lambda_i \neq 0 \quad , \quad r \geq 1 \quad , \quad r + s \leq \dim(\Delta_R)$$

$$\sum_{i=1}^r \lambda_i \begin{bmatrix} g_0(x_i) \\ \vdots \\ g_m(x_i) \\ -R(x_i)h_0(x_i) \\ \vdots \\ -R(x_i)h_n(x_i) \end{bmatrix} + \sum_{i=1}^s \mu_i \begin{bmatrix} \alpha_0(t_i) \\ \vdots \\ \alpha_m(t_i) \\ \beta_0(t_i) \\ \vdots \\ \beta_n(t_i) \end{bmatrix} + \sum_{i=1}^K \gamma_i \begin{bmatrix} \alpha_{i0} \\ \vdots \\ \alpha_{im} \\ \beta_{i0} \\ \vdots \\ \beta_{in} \end{bmatrix} = 0$$

1°) Soit $(f, c) \in \mathcal{B}$ et soit $R(x)$ optimal pour f dans la classe $W_{m,n}$ associée à $c(t)$. En supposant que R possède la propriété (P) ci-dessus, nous allons montrer que R est le seul élément de $W_{m,n}$ optimal pour f .

Supposons au contraire qu'il existe dans $W_{m,n}$ un autre élément optimal $R'(x)$ et posons :

$$R(x) = \frac{\sum_{j=0}^m a_j^* g_j(x)}{\sum_{j=0}^n b_j^* h_j(x)}, \quad R'(x) = \frac{\sum_{j=0}^m a_j' g_j(x)}{\sum_{j=0}^n b_j' h_j(x)}$$

$R(x)$ étant optimal, il existe, d'après la remarque 3 du théorème 5, des (e) -points de $f - R$, des points $t_1, \dots, t_s$ de G et des nombres $\lambda_1, \dots, \lambda_r, \mu_1, \dots, \mu_s, \gamma_1, \dots, \gamma_K$ tels que :

$$(35) \quad \begin{cases} \lambda_i y_R(x_i) > 0, \quad \mu_i > 0 \\ r \geq 1, \quad r + s \leq 1 + \dim(\Delta_R) \\ \sum_{j=0}^m a_j^* \alpha_j(t_i) + \sum_{j=0}^n b_j^* \beta_j(t_i) = c(t_i) \quad (i = 1, \dots, s) \end{cases}$$

$$(36) \quad \sum_{i=1}^r \lambda_i \begin{bmatrix} g_0(x_i) \\ \vdots \\ g_m(x_i) \\ -R(x_i)h_0(x_i) \\ \vdots \\ -R(x_i)h_n(x_i) \end{bmatrix} + \sum_{i=1}^s \mu_i \begin{bmatrix} \alpha_0(t_i) \\ \vdots \\ \alpha_m(t_i) \\ \beta_0(t_i) \\ \vdots \\ \beta_n(t_i) \end{bmatrix} + \sum_{i=1}^K \gamma_i \begin{bmatrix} \alpha_{i0} \\ \vdots \\ \alpha_{im} \\ \beta_{i0} \\ \vdots \\ \beta_{in} \end{bmatrix} = 0$$

De plus, puisque $R(x)$ possède la propriété (P), on a exactement :

$$(37) \quad r + s = 1 + \dim(\Delta_R)$$

Considérons dans $\mathcal{D}_R$ l'élément suivant :

$$\Phi'(x) = \sum_{j=0}^m a'_j g_j(x) - R(x) \sum_{j=0}^n b'_j h_j(x)$$

ou encore :

$$\Phi'(x) = \{[f(x) - R(x)] - [f(x) - R'(x)]\} \sum_0^n b'_j h_j(x)$$

Mais on a :

$$|f(x) - R'(x)| \leq \|f - R'\| = \|f - R\| = |f(x) - R(x)| \quad \forall x \in E_R$$

Et donc :

$$\Phi'(x) [f(x) - R(x)] \geq 0 \quad \forall x \in E_R$$

ou encore, en tenant compte de (35) :

$$(38) \quad \lambda_i \Phi'(x_i) \geq 0 \quad i = 1, \dots, r$$

Mais, on a aussi (voir corollaire 1 du théorème 5) :

$$(39) \quad \sum_{i=1}^r \lambda_i \Phi'(x_i) + \sum_{i=1}^s \mu_i \left[\sum_{j=0}^m a'_j \alpha_j(t_i) + \sum_{j=0}^n b'_j \beta_j(t_i) - c(t_i) \right] = 0$$

Comme $R' \in W_{m,n}$ et $\mu_i > 0$, des relations (38) et (39), on déduit :

$$(40) \quad \begin{cases} \Phi'(x_i) = 0 & i = 1, \dots, r \\ \sum_{j=0}^m a'_j \alpha_j(t_i) + \sum_{j=0}^n b'_j \beta_j(t_i) = c(t_i) & i = 1, \dots, s \end{cases}$$

Comme au paragraphe 3, introduisons une base $\Phi_1(x), \dots, \Phi_M(x)$ de Δ_R ainsi que les fonctions $\phi_i(t)$ définies par (11) et (12) et les fonctions $\Phi^*(x)$ et $\phi^*(t)$ définies par (13).

Les formules (40) s'écrivent alors :

$$\begin{cases} \sum_{j=1}^M \tau'_j \Phi_j(x_i) = 0 & i = 1, \dots, r \\ \sum_{j=1}^M \tau'_j \phi_j(t_i) = c(t_i) - \phi^*(t_i) & i = 1, \dots, s \end{cases}$$

ou encore, en tenant compte de (35) :

$$(41) \quad \begin{cases} \sum_{j=1}^M \tau_j' \Phi_j(x_i) = 0 & i = 1, \dots, r \\ \sum_{j=1}^M \tau_j' \Phi_j(t_i) = 0 & i = 1, \dots, s \end{cases}$$

La fonction $\Phi'(x)$ n'étant pas identiquement nulle, (41) est un système linéaire homogène de $r + s = M + 1$ équations à M inconnues admettant une solution non triviale. Ce qui prouve que l'on a :

$$\begin{vmatrix} \Phi_1(x_2) & \dots & \Phi_M(x_2) \\ \dots & \dots & \dots \\ \Phi_1(x_r) & \dots & \Phi_M(x_r) \\ \Phi_1(t_1) & \dots & \Phi_M(t_1) \\ \dots & \dots & \dots \\ \Phi_1(t_s) & \dots & \Phi_M(t_s) \end{vmatrix} = 0$$

ce qu'on peut encore écrire sous la forme :

$$(42) \quad \sum_{i=2}^r \lambda_i' \begin{bmatrix} \Phi_1(x_i) \\ \vdots \\ \Phi_M(x_i) \end{bmatrix} + \sum_{i=1}^s \mu_i' \begin{bmatrix} \Phi_1(t_i) \\ \vdots \\ \Phi_M(t_i) \end{bmatrix} = 0$$

avec des coefficients λ_i', μ_i' non tous nuls.

Dans la formule (42), on peut supposer que l'on a :

$$\begin{cases} \mu_i' \leq \mu_i & i \neq i_0 \\ \mu_{i_0}' = \mu_{i_0} \end{cases}$$

La formule (36) s'écrit aussi :

$$(43) \quad \sum_{i=1}^r \lambda_i \begin{bmatrix} \Phi_1(x_i) \\ \vdots \\ \Phi_M(x_i) \end{bmatrix} + \sum_{i=1}^s \mu_i \begin{bmatrix} \Phi_1(t_i) \\ \vdots \\ \Phi_M(t_i) \end{bmatrix} = 0$$

En retranchant (42) de (43), on obtient :

$$\lambda_1 \begin{bmatrix} \Phi_1(x_1) \\ \vdots \\ \Phi_M(x_1) \end{bmatrix} + \sum_{i=2}^r (\lambda_i - \lambda'_i) \begin{bmatrix} \Phi_1(x_i) \\ \vdots \\ \Phi_M(x_i) \end{bmatrix} + \sum_{\substack{i=1 \\ i \neq i_0}}^s (\mu_i - \mu'_i) \begin{bmatrix} \Phi_1(t_i) \\ \vdots \\ \Phi_M(t_i) \end{bmatrix} = 0$$

On en déduit comme au paragraphe précédent, que l'on a :

$$\sum_{i=1}^{r'} \lambda''_i \begin{bmatrix} g_0(x_i) \\ \vdots \\ -R(x_i)h_n(x_i) \end{bmatrix} + \sum_{i=1}^{s'} \mu''_i \begin{bmatrix} \alpha_0(t_i) \\ \vdots \\ \beta_n(t_i) \end{bmatrix} + \sum_{i=1}^K \gamma''_i \begin{bmatrix} \alpha_{i0} \\ \vdots \\ \beta_{in} \end{bmatrix} = 0$$

avec :

$$\lambda''_i \neq 0, \mu''_i > 0, r' \geq 1, r' + s' \leq M = \dim(\Delta_R)$$

Ceci est en contradiction avec le fait que $R(x)$ possède la propriété (P).
Donc $R'(x)$ ne peut exister et $R(x)$ est bien le seul élément optimal pour f .

2°) Réciproquement nous allons montrer que s'il existe une fraction $R(x)$ ne possédant pas la propriété (P), on peut construire un élément $(f, c) \in \mathcal{B}$ tel que $f(x)$ possède deux éléments optimaux distincts dans la classe $W_{m,n}$ associée à la fonction $c(t)$.

Supposons donc qu'il existe une fraction $R(x) = \sum_{j=0}^m a_j^* g_j(x) / \sum_{j=0}^n b_j^* h_j(x)$,
des points $x_1, \dots, x_r$ de X , des points $t_1, \dots, t_s$ de G et des nombres $\lambda_i, \mu_i, \gamma_i$ tels que :

$$(44) \left\{ \begin{array}{l} \sum_{j=0}^m \alpha_{ij} a_j^* + \sum_{j=0}^n \beta_{ij} b_j^* = c_i, \quad i = 1, \dots, K \\ \sum_{j=0}^n b_j h_j(x) > 0 \quad \text{sur } X \\ \mu_i > 0, \lambda_i \neq 0, r \geq 1, r + s \leq \dim(\Delta_R) \\ \sum_{i=1}^r \lambda_i \begin{bmatrix} g_0(x_i) \\ \vdots \\ -R(x_i)h_n(x_i) \end{bmatrix} + \sum_{i=1}^s \mu_i \begin{bmatrix} \alpha_0(t_i) \\ \vdots \\ \beta_n(t_i) \end{bmatrix} + \sum_{i=1}^K \gamma_i \begin{bmatrix} \alpha_{i0} \\ \vdots \\ \beta_{in} \end{bmatrix} = 0 \end{array} \right.$$

(a) Montrons d'abord qu'il existe dans Δ_R un élément non identiquement nul :

$$\Phi(x) \equiv \sum_0^m a_j g_j(x) - R(x) \sum_0^n b_j h_j(x)$$

tel que :

$$(45) \quad \left\{ \begin{array}{ll} \Phi(x_i) = 0 & i = 1, \dots, r \\ \sum_{j=0}^m (a_j - a_j^*) \alpha_j(t_i) + \sum_{j=0}^n (b_j - b_j^*) \beta_j(t_i) = 0 & i = 1, \dots, s \end{array} \right.$$

En reprenant les notations utilisées dans la démonstration du théorème 5 (formules (10) à (13)), les conditions ci-dessus s'écrivent :

$$\left\{ \begin{array}{ll} \Phi(x_i) \equiv \sum_{j=1}^M \tau_j \Phi_j(x_i) = 0 & i = 1, \dots, r \\ \varphi(t_i) - \varphi^*(t_i) \equiv \sum_{j=1}^M \tau_j \varphi_j(t_i) = 0 & i = 1, \dots, s \end{array} \right.$$

$\Phi_1(x), \dots, \Phi_M(x)$ étant une base de Δ_R , il faut donc montrer qu'il existe une solution non triviale pour le système linéaire homogène suivant :

$$(46) \quad \left\{ \begin{array}{ll} \sum_{j=1}^M \tau_j \Phi_j(x_i) = 0 & i = 1, \dots, r \\ \sum_{j=1}^M \tau_j \varphi_j(t_i) = 0 & i = 1, \dots, s \end{array} \right.$$

Or, d'après (44), on a :

$$\sum_{i=1}^r \lambda_i \Phi(x_i) + \sum_{i=1}^s \mu_i \varphi(t_i) = 0 \quad \forall \Phi \in \Delta_R$$

ou encore :

$$(47) \quad \sum_{i=1}^r \lambda_i \begin{bmatrix} \Phi_1(x_i) \\ \vdots \\ \Phi_M(x_i) \end{bmatrix} + \sum_{i=1}^s \mu_i \begin{bmatrix} \varphi_1(t_i) \\ \vdots \\ \varphi_M(t_i) \end{bmatrix} = 0$$

En supposant que, dans (44), $r + s$ est minimum, ceci montre que le système (46) est exactement de rang $r + s - 1 \leq M - 1$.

Comme c'est un système homogène à M inconnues, il y a donc une solution non triviale.

(b) Désignons donc par $\Phi(x)$ un élément de Δ_R non identiquement nul vérifiant (45) :

$$\Phi(x) = \sum_{j=0}^m a_j g_j(x) - R(x) \sum_{j=0}^n b_j h_j(x)$$

Considérons alors la fraction suivante :

$$R_1(x) = \frac{\sum_{j=0}^m a'_j g_j(x)}{\sum_{j=0}^n b'_j h_j(x)}$$

avec :

$$\begin{cases} a'_j = \theta a_j + (1 - \theta) a_j^* & j = 0, 1, \dots, m \\ b'_j = \theta b_j + (1 - \theta) b_j^* & j = 0, 1, \dots, n \\ 0 < \theta < 1 \end{cases}$$

Puisque $\sum_{j=0}^n b_j^* h_j(x) > 0$ sur X , nous pouvons choisir θ assez petit pour que :

$$\sum_{j=0}^n b'_j h_j(x) > 0 \quad \text{sur } X$$

De plus, les nombres a'_j , b'_j vérifient évidemment :

$$\sum_{j=0}^m \alpha_{ij} a'_j + \sum_{j=0}^n \beta_{ij} b'_j = c_i \quad i = 1, \dots, K$$

(c) Définissons maintenant une fonction $c(t) \in C[G]$ vérifiant la condition de régularité.

Pour cela, montrons tout d'abord qu'il existe des nombres $\bar{a}_j, \bar{b}_j$ tels que :

$$(48) \quad \begin{cases} \sum_{j=0}^m \alpha_{ij} \bar{a}_j + \sum_{j=0}^n \beta_{ij} \bar{b}_j = c_i & i = 1, \dots, K \\ \sum_{j=0}^m (\bar{a}_j - a_j^*) \alpha_j(t_i) + \sum_{j=0}^n (\bar{b}_j - b_j^*) \beta_j(t_i) > 0 & i = 1, \dots, s \end{cases}$$

Cela revient (en reprenant les notations précédemment utilisées) à montrer qu'il existe des nombres $\bar{\tau}_j$ tels que :

$$(49) \quad \sum_{j=1}^M \bar{\tau}_j \varphi_j(t_i) > 0 \quad i = 1, \dots, s$$

La matrice du système (49) est de rang s (on suppose $s \geq 1$ sinon le problème de l'existence d'une solution de (48) ne se pose pas). Sinon on aurait :

$$(50) \quad \sum_{i=1}^s \mu'_i \begin{bmatrix} \varphi_1(t_i) \\ \vdots \\ \varphi_M(t_i) \end{bmatrix} = 0$$

avec des coefficients μ'_i non tous nuls. On pourrait choisir ces coefficients tels que :

$$\begin{cases} \mu'_i \leq \mu_i & i = 1, \dots, s \\ \mu'_{i0} = \mu_{i0} \end{cases}$$

En retranchant (50) de (47), on obtiendrait alors :

$$\sum_{i=1}^r \lambda_i \begin{bmatrix} \Phi_1(x_i) \\ \vdots \\ \Phi_M(x_i) \end{bmatrix} + \sum_{\substack{i=1 \\ i \neq i_0}}^s (\mu_i - \mu'_i) \begin{bmatrix} \varphi_1(t_i) \\ \vdots \\ \varphi_M(t_i) \end{bmatrix} = 0$$

ce qui est impossible si l'on suppose que, dans la formule (44), $r + s$ est minimum.

La matrice du système (49) est donc de rang $s \leq M - r \leq M - 1$ ce qui prouve qu'il existe une solution $(\bar{a}_0, \dots, \bar{a}_m, \bar{b}_0, \dots, \bar{b}_n)$ de (48).

Nous choisirons pour $c(t)$ une fonction de $\mathcal{O}[G]$ vérifiant les conditions suivantes :

$$(51) \quad \left\{ \begin{array}{ll} c(t) \leq \sum_0^m a_j^* \alpha_j(t) + \sum_0^n b_j^* \beta_j(t) & \forall t \in G \\ c(t) \leq \sum_0^m a_j' \alpha_j(t) + \sum_0^n b_j' \beta_j(t) & \forall t \in G \\ c(t_i) = \sum_0^m a_j^* \alpha_j(t_i) + \sum_0^n b_j^* \beta_j(t_i) & i = 1, \dots, s \\ c(t) < \sum_0^m \bar{a}_j \alpha_j(t) + \sum_0^n \bar{b}_j \beta_j(t) & \forall t \in G - \{t_1, \dots, t_s\} \end{array} \right.$$

D'après (48) et (51), on a :

$$\begin{cases} \sum_0^m \alpha_{ij} \bar{a}_j + \sum_0^n \beta_{ij} \bar{b}_j = c_i & i = 1, \dots, K \\ \sum_0^m \bar{a}_j \alpha_j(t) + \sum_0^n \bar{b}_j \beta_j(t) > c(t) & \forall t \in G \end{cases}$$

ce qui prouve que la fonction $c(t)$ vérifie la condition de régularité.

(d) D'après (b) et les formules (44), (51), on voit que $R(x)$ et $R_1(x)$ sont deux éléments de la classe $W_{m,n}$ associée à la fonction $c(t)$. De plus, on a :

$$R(x) - R_1(x) = \frac{\sum_0^m a_j^* g_j(x)}{\sum_0^n b_j^* h_j(x)} - \frac{\sum_0^m [\theta a_j + (1 - \theta) a_j^*] g_j(x)}{\sum_0^n [\theta b_j + (1 - \theta) b_j^*] h_j(x)}$$

ou :

$$(52) \quad R(x) - R_1(x) = \frac{-\theta \Phi(x)}{\sum_0^n [\theta b_j + (1 - \theta) b_j^*] h_j(x)}$$

Ceci montre que R et R_1 sont deux éléments distincts de $W_{m,n}$ qui vérifient en outre :

$$R(x_i) = R_1(x_i) \quad i = 1, \dots, r$$

et par conséquent la fraction $R_1(x)$ vérifie les relations (44). Enfin, en choisissant θ assez petit, on pourra supposer que :

$$\|R - R_1\| \leq 1$$

(e) Choisissons maintenant une fonction $f \in C[X]$ pour laquelle R et R_1 sont optimaux dans $W_{m,n}$.

Nous prendrons :

$$f(x) = R(x) + [1 - |R(x) - R_1(x)|] g(x)$$

avec :

$$\begin{cases} g \in C[X] & , \quad \|g\| = 1 \\ g(x_i) = \text{sign}(\lambda_i) & i = 1, \dots, r \end{cases}$$

On a immédiatement :

$$(53) \quad f(x_i) - R(x_i) = f(x_i) - R_1(x_i) = \text{sign}(\lambda_i) \quad i = 1, \dots, r$$

$$|y_R(x)| \leq [1 - |R(x) - R_1(x)|] \|g\| \leq 1$$

$$|y_{R_1}(x)| \leq |R(x) - R_1(x)| + [1 - |R(x) - R_1(x)|] = 1$$

Donc :

$$\|f - R\| = \|f - R_1\| = 1$$

La formule (53) montre que $x_1, \dots, x_r$ sont des (e)-points de $f - R$ et $f - R_1$ tels que :

$$\lambda_i [f(x_i) - R(x_i)] = \lambda_i [f(x_i) - R_1(x_i)] > 0 \quad i = 1, \dots, r$$

De plus, on a :

$$\begin{cases} \sum_{j=0}^m a_j^* \alpha_j(t_i) + \sum_{j=0}^n b_j^* \beta_j(t_i) = c(t_i) & i = 1, \dots, s \\ \sum_{j=0}^m a'_j \alpha_j(t_i) + \sum_{j=0}^n b'_j \beta_j(t_i) = c(t_i) & i = 1, \dots, s \end{cases}$$

Comme les fractions $R(x)$ et $R_1(x)$ vérifient la relation (44), ceci prouve (remarque 3 du théorème 5) que $f(x)$ possède dans $W_{m,n}$ deux éléments optimaux distincts $R(x)$ et $R_1(x)$.

Donc, s'il y a unicité de la meilleure approximation pour tout élément $(f, c) \in \mathcal{B}$, toute fraction $R(x)$ possède la propriété (P).

Remarque 1 : La première partie de la démonstration du théorème 6 fournit une condition suffisante d'unicité qu'on peut énoncer sous la forme suivante :

Théorème 7 : soit $R \in W_{m,n}$ optimal pour f . Si R possède la propriété (P), alors R est le seul élément optimal pour f dans $W_{m,n}$.

Remarque 2 : Comme on l'a déjà fait dans la remarque 4 du théorème 5, on peut prendre les contraintes sous la forme :

$$\begin{cases} \sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i & i = 1, \dots, K \\ \left| \sum_{j=0}^m a_j \alpha_j(t) + \sum_{j=0}^n b_j \beta_j(t) - \Gamma(t) \right| \leq M & \forall t \in G \end{cases}$$

Dans la propriété (P) , on peut alors se contenter de prendre $\mu_i \neq 0$ (au lieu de $\mu_i > 0$). La démonstration précédente reste valable avec quelques très légères modifications en remarquant que $\mu_i > 0$ correspond à :

$$\sum_0^m a_j \alpha_j(t_i) + \sum_0^n b_j \beta_j(t_i) = \Gamma(t_i) - M$$

et que $\mu_i < 0$ correspond à :

$$\sum_0^m a_j \alpha_j(t_i) + \sum_0^n b_j \beta_j(t_i) = \Gamma(t_i) + M$$

Lorsque $K = 0$ et $n = 0$, le théorème 7 nous redonne le résultat démontré dans le cas polynomial par TERRENOIRE ([1967], proposition 7-2).

Remarque 3 : dans le cas présent, le théorème I-6 et son corollaire sur l'unicité ne s'appliquent pas car l'ensemble Ω n'est pas strictement convexe.

Remarque 4 : nous donnerons au chapitre suivant (paragraphe 8) une autre forme du théorème 7 dans le cas où $X = [a, b]$.

5 - Propriétés d'alternance

On a seulement les théorèmes I - 7 et I - 8 , c'est-à-dire :

(A) soit $R \in W_{m,n}$. S'il existe des points $x_1 < \dots < x_N$ de $[a, b] = X$ tels que :

$$[f(x_{i+1}) - R(x_{i+1})] [f(x_i) - R(x_i)] \leq 0$$

$$N \geq 1 + \nu(\mathcal{D}_R)$$

alors :

$$\rho \geq \min_{1 \leq i \leq N} |f(x_i) - R(x_i)|$$

(B) Si $f - R$ possède au moins $1 + \nu(\mathcal{D}_R)$ (e)-points alternés, alors R est optimal pour f .

Remarquons que l'on a :

$$\mathcal{D}_R \subseteq \Delta_R$$

ce qui entraîne :

$$\nu(\mathcal{D}_R) \leq \nu(\Delta_R)$$

Les résultats (A) et (B) ci-dessus restent donc valables en remplaçant $\nu(\mathcal{D}_R)$ par $\nu(\Delta_R)$, qui peut dans certains cas être plus facile à déterminer.

Nous verrons au chapitre suivant (paragraphe 8) d'autres propriétés d'alternance.

Chapitre III

APPROXIMATION RATIONNELLE AVEC CONTRAINTES DU TYPE EGALITE

Dans ce chapitre, on étudie l'approximation d'une fonction continue $f(x)$ par des fractions rationnelles généralisées :

$$R(x) = \frac{\sum_0^m a_j g_j(x)}{\sum_0^n b_j h_j(x)}$$

dont les coefficients vérifient des égalités linéaires :

$$\sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i \quad i = 1, \dots, K$$

Le problème est posé au paragraphe 1. Au paragraphe 2, on examine ce que deviennent dans ce cas particulier, les théorèmes sur l'existence, la convexité de l'ensemble des éléments optimaux et le théorème de La Vallée-Poussin. Au paragraphe 3, on énonce plusieurs formes équivalentes du théorème de caractérisation des éléments optimaux. Les propriétés d'alternance sont étudiées au paragraphe 4. Dans les paragraphes 5 et 6, on étudie l'unicité et "l'unicité forte". Dans le paragraphe 7, on examine la continuité de l'opérateur qui à une fonction $f(x)$ associe l'ensemble de ses éléments optimaux. Enfin dans le paragraphe 8, on utilise les résultats du présent chapitre pour compléter les théorèmes énoncés au chapitre précédent.

1 - Introduction.

On considère encore un espace compact X et $C[X]$ espace vectoriel des fonctions continues sur X , à valeurs réelles. On se donne dans $C[X]$ des fonctions $g_0(x), \dots, g_m(x), h_0(x), \dots, h_n(x)$ et on suppose les $g_i(x)$ d'une part, les $h_i(x)$ d'autre part linéairement indépendantes sur X .

On désigne par $W_{m,n}$ la classe des fonctions :

$$R(x) = \frac{\sum_0^m a_j g_j(x)}{\sum_0^n b_j h_j(x)}$$

où les a_j, b_j sont des nombres réels tels que :

$$(1) \quad \sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i \quad i = 1, \dots, K \quad (K \geq 1)$$

$$(2) \quad \sum_{j=0}^n b_j h_j(x) > 0 \quad \text{sur } X$$

Les nombres α_{ij}, β_{ij} et c_i sont donnés. On suppose que les vecteurs $V_i (i=1, \dots, K)$ de composantes $(\alpha_{i0}, \alpha_{i1}, \dots, \alpha_{im}, \beta_{i0}, \beta_{i1}, \dots, \beta_{in})$ sont linéairement indépendants.

Etant donné $f \in C[X]$, $f \notin W_{m,n}$, on cherche dans $W_{m,n}$ un élément R^* tel que :

$$\|f - R^*\| \leq \|f - R\| \quad \forall R \in W_{m,n}$$

Comme au chapitre précédent, nous utiliserons les notations suivantes :

$$y_R(x) = f(x) - R(x)$$

$$\rho = \inf_{R \in W_{m,n}} \|f - R\|$$

$$E_R = \{x \in X \mid f(x) - R(x) = \pm \|f - R\|\}$$

Enfin les classes de fonctions $\mathcal{D}_R$ et Δ_R définies au chapitre précédent sont ici confondues. C'est l'ensemble des fonctions :

$$\Phi(x) = \sum_{j=0}^m a_j' g_j(x) - R(x) \sum_{j=0}^n b_j' h_j(x)$$

où les a_j', b_j' sont des nombres réels vérifiant (1).

2 - Existence d'un élément optimal.

Désignons encore par $\overline{W}_{m,n}$ la classe des fonctions $R \in C[X]$ qui vérifient une identité de la forme :

$$R(x) \sum_{i=0}^n b_i h_i(x) \equiv \sum_{i=0}^m a_i g_i(x)$$

où les a_i, b_i sont des nombres réels non tous nuls vérifiant la condition (1).

Les théorèmes II-1, II-2 et II-3 du chapitre précédent nous donnent immédiatement les résultats suivants :

Théorème 1 : si $c_i = 0$ ($i = 1, \dots, K$) et si les fonctions $h_0(x), h_1(x), \dots, h_n(x)$ sont linéairement indépendantes sur tout sous-ensemble ouvert de X , alors toute fonction $f \in C[X]$ possède dans $\overline{W}_{m,n}$ au moins un élément optimal.

Théorème 2 : l'ensemble des points de R^{m+n+2} dont les coordonnées $(a_0, a_1, \dots, a_m, b_0, b_1, \dots, b_n)$ sont les coefficients des éléments $R \in W_{m,n}$ optimaux pour f est convexe.

Théorème 3 : soit $R \in W_{m,n}$. S'il existe un ensemble D de points de X tel que :

- a) $f(x) - R(x) \neq 0 \quad \forall x \in D,$
- b) $\inf_{x \in D} \Phi(x) [f(x) - R(x)] \leq 0 \quad \forall \Phi \in \mathcal{D}_R$

alors, on a :

$$\rho \geq \inf_{x \in D} |f(x) - R(x)|$$

Bien entendu, ce théorème peut aussi se mettre sous la forme indiquée dans la remarque 2 du théorème I-3.

3- Caractérisation des éléments optimaux.

Le théorème II-4 nous donne tout d'abord une caractérisation analogue à celle démontrée par BROSOWSKI [1965] dans le cas de l'approximation par des polynômes généralisés dont les coefficients vérifient des égalités linéaires :

Théorème 4 : pour que $R \in W_{m,n}$ soit optimal pour f , il faut et il suffit que l'on ait :

$$\inf_{x \in E_R} \Phi(x) [f(x) - R(x)] \leq 0 \quad \forall \Phi \in \mathcal{D}_R$$

Dans le cas présent, $\mathcal{D}_R$ est un sous-espace de $C[X]$. Désignons par $\Phi_1(x), \dots, \Phi_N(x)$ une base de $\mathcal{D}_R$. La relation ci-dessus signifie que le système linéaire suivant, par rapport aux inconnues $\tau_1, \dots, \tau_N$, est incompatible :

$$\sum_{i=1}^N \tau_i \Phi_i(x) [f(x) - R(x)] > 0 \quad \forall x \in E_R$$

D'après le théorème 0-2, on peut donc énoncer le théorème 4 sous la forme suivante :

Théorème 5 : pour que $R \in W_{m,n}$ soit optimal pour f , il faut et il suffit que l'origine de R^N appartienne à l'enveloppe convexe de l'ensemble des points :

$$\{[y_R(x)\Phi_1(x), \dots, y_R(x)\Phi_N(x)] \in R^N \mid x \in E_R\},$$

où $\Phi_1(x), \dots, \Phi_N(x)$ désigne une base arbitraire de $\mathcal{D}_R$.

Le corollaire 1 du théorème II-5 s'énonce simplement sous la forme suivante :

Théorème 6 : pour que $R \in W_{m,n}$ soit optimal pour f , il faut et il suffit qu'il existe des (e) -points $x_1, \dots, x_r$ de $f - R$ et des nombres

$\lambda_1, \dots, \lambda_r$ tels que :

$$\begin{cases} 1 \leq r \leq 1 + \dim(\mathcal{D}_R) \\ \lambda_i [f(x_i) - R(x_i)] > 0 & i = 1, \dots, r \\ \sum_{i=1}^r \lambda_i \Phi(x_i) = 0 & \forall \Phi \in \mathcal{D}_R \end{cases}$$

On retrouve avec les théorèmes 4, 5 et 6, les trois formes de caractérisation des éléments optimaux données par CHENEY [1966] dans le cas de l'approximation par fractions rationnelles généralisées sans contraintes.

Enfin, dans le cas présent, le corollaire 3 du théorème II-5 nous permet d'énoncer :

Théorème 7 : pour que $R \in W_{m,n}$ soit optimal pour f , il faut et il suffit qu'il existe une fonctionnelle linéaire L telle que :

$$\begin{aligned} L(f-R) &= \|f-R\|, & \|L\| &= 1 \\ L(\Phi) &= 0 & \forall \Phi \in \mathcal{D}_R \quad (\text{voir p. 149}) \end{aligned}$$

On retrouve la forme du théorème de caractérisation donnée par SINGER [1956, 1959] dans le cas linéaire avec ou sans contraintes.

4- Propriétés d'alternance.

Dans ce paragraphe nous supposerons que $X=[a,b]$ intervalle fini de la droite réelle.

Nous dirons que $f-R$ présente N (ε)-points alternés s'il existe N points $x_1 < \dots < x_N$ de $[a,b]$ tels que :

$$f(x_i) - R(x_i) = (-1)^i \varepsilon \|f-R\| \quad \text{avec} \quad \varepsilon = \pm 1.$$

D'autre part, nous dirons qu'un sous-espace M de dimension finie de $C[a,b]$ est un sous-espace de Haar si tout élément $\Phi \in M$ non identiquement nul a dans $[a,b]$ au plus $\dim(M) - 1$ zéros. Rappelons que ceci est

équivalent, en désignant par $\Phi_1(x), \dots, \Phi_N(x)$ une base de M , à la relation :

$$\begin{vmatrix} \Phi_1(x_1) & \dots & \Phi_1(x_N) \\ \dots & \dots & \dots \\ \Phi_N(x_1) & \dots & \Phi_N(x_N) \end{vmatrix} \neq 0$$

quels que soient les points $x_1, \dots, x_N$ distincts dans $[a, b]$.

M étant un sous-espace de dimension finie de $C[a, b]$, nous poserons (comme au chapitre I) :

$$\nu(M) = 1 + \text{nombre maximum de changements de signe dans } [a, b] \text{ des éléments de } M$$

et en outre :

$$\eta(M) = \max \{ \dim(H) \text{ , où } H \text{ est un sous-espace de Haar de } M \}$$

On a évidemment :

$$\nu(M) \geq \dim(M) \geq \eta(M)$$

Les théorèmes I-7 et I-8 nous donnent tout d'abord les résultats suivants :

Théorème 8 : soit $R \in W_{m,n}$. S'il existe des points $x_1 < \dots < x_N$ de $[a, b]$ tels que :

$$[f(x_{i+1}) - R(x_{i+1})][f(x_i) - R(x_i)] < 0 \quad i = 1, \dots, N-1$$

$$N \geq 1 + \nu(\mathcal{D}_R)$$

alors, on a :

$$\rho \geq \min_{1 \leq i \leq N} |f(x_i) - R(x_i)|$$

Théorème 9 : si $f - R$ possède au moins $[1 + \nu(\mathcal{D}_R)]$ (e)- points alternés, alors R est optimal pour f dans $W_{m,n}$.

Inversement, nous allons démontrer un résultat qui généralise celui donné par CHENEY [1966] dans le cas de l'approximation par fractions rationnelles généralisées sans contraintes sur les coefficients :

Théorème 10 : si R est optimal pour f , le nombre de (e) -points alternés de $f-R$ est au moins égal à $1 + \eta(\mathcal{D}_R)$.

Soit M un sous-espace de Haar de $\mathcal{D}_R$ de dimension $\eta(\mathcal{D}_R)$.

Posons $N = \eta(\mathcal{D}_R)$ et désignons par $\Phi_1(x), \dots, \Phi_N(x)$ une base de M .

$R(x)$ étant optimal pour f , il existe, d'après le théorème 6, des (e) -points $x_1, \dots, x_r$ de $f-R$ et des nombres $\lambda_1, \dots, \lambda_r$ tels que :

$$\begin{cases} \lambda_i y_R(x_i) > 0 & , 1 \leq i \leq r \\ \sum_{i=1}^r \lambda_i \Phi(x_i) = 0 & \forall \Phi \in \mathcal{D}_R \end{cases}$$

On a donc en particulier :

$$(3) \quad \lambda_1 \Phi_j(x_1) + \dots + \lambda_r \Phi_j(x_r) = 0 \quad j = 1, \dots, N$$

On a alors :

$$(4) \quad r = N + 1$$

Sinon en supposant $r \leq N$, on construit à partir de la suite $x_1, \dots, x_r$ une suite $x_1, \dots, x_N$ de N points de $[a, b]$, deux à deux distincts et on considère le système linéaire homogène suivant :

$$(5) \quad \tau_1 \Phi_j(x_1) + \dots + \tau_N \Phi_j(x_N) = 0 \quad j = 1, \dots, N$$

D'après (3), ce système admet la solution :

$$\begin{cases} \tau_i = \lambda_i & 1 \leq i \leq r \\ \tau_i = 0 & r < i \leq N \end{cases}$$

Mais, M étant un sous-espace de Haar, le déterminant du système (5) est non nul, c'est-à-dire que ce système n'admet pas de solution non nulle. Il y a contradiction, ce qui prouve que $r > N$ et donc que $r = N + 1$.

On a donc, en posant :

$$\lambda_i = \theta_i y_R(x_i)$$

le résultat suivant :

$$\sum_{i=1}^{N+1} \theta_i y_R(x_i) \Phi_j(x_i) = 0 \quad j = 1, \dots, N$$

avec :

$$\theta_i > 0$$

Ce qui prouve que l'origine de R^N appartient à l'enveloppe convexe de l'ensemble des points :

$$\{y_R(x_i) [\Phi_1(x_i), \dots, \Phi_N(x_i)] \mid i = 1, \dots, N+1\}$$

Supposons que l'on ait $x_1 < \dots < x_{N+1}$.

D'après le théorème 0-4, on a :

$$y_R(x_i) y_R(x_{i+1}) < 0 \quad i = 1, \dots, N$$

Ce qui montre que le nombre de (e) -points alternés de $f-R$ est au moins égal à $N+1 = 1 + \eta(\mathcal{D}_R)$.

Remarque 1 : si $\mathcal{D}_R$ est un sous-espace de Haar, on a alors :

$$\nu(\mathcal{D}_R) = \dim(\mathcal{D}_R) = \eta(\mathcal{D}_R).$$

Des théorèmes 9 et 10, on déduit donc le résultat particulier suivant :

Corollaire : supposons que $\mathcal{D}_R$ soit un sous-espace de Haar pour tout $R \in W_{m,n}$. Pour que $R_0 \in W_{m,n}$ soit optimal pour f , il faut et il suffit que le nombre de (e) -points alternés de $f-R_0$ soit au moins égal à $1 + \dim(\mathcal{D}_{R_0})$.

Ce résultat peut aussi s'obtenir en appliquant les résultats généraux sur l'approximation non linéaire (voir MEINARDUS [1964], théorème 86).

5- Unicité.

Nous allons tout d'abord donner un résultat déduit du chapitre précédent, valable pour tout X compact.

Dans le cas où $X = [a, b]$ intervalle fini de la droite réelle, nous démontrerons une autre condition nécessaire et suffisante d'unicité qui généralise celle donnée dans le cas de l'approximation sans contraintes. Nous verrons ensuite l'équivalence entre les deux résultats donnés.

Désignons par $\tilde{C}[X]$ l'ensemble des fonctions $f \in C[X]$ qui possèdent au moins un élément optimal dans $W_{m,n}$.

D'après le théorème II-6, nous avons le résultat suivant :

Théorème 11 : toute fonction $f \in \tilde{C}[X]$ a dans $W_{m,n}$ un élément optimal unique si et seulement si toute fraction $R \in W_{m,n}$ possède la propriété (P) ci-dessous.

Propriété (P) : il n'existe pas de points $x_1, \dots, x_r$ de X et de nombres λ_i, γ_i tels que :

$$\lambda_i \neq 0, \quad 1 \leq r \leq \dim(\mathcal{D}_R)$$

$$(6) \quad \sum_{i=1}^r \lambda_i \begin{bmatrix} g_0(x_i) \\ \vdots \\ g_m(x_i) \\ -R(x_i)h_0(x_i) \\ \vdots \\ -R(x_i)h_n(x_i) \end{bmatrix} + \sum_{i=1}^K \gamma_i \begin{bmatrix} \alpha_{i0} \\ \vdots \\ \alpha_{im} \\ \beta_{i0} \\ \vdots \\ \beta_{in} \end{bmatrix} = 0$$

•

Remarque : comme on l'a déjà vu dans la démonstration du théorème II-5, la condition (6) ci-dessus est équivalente à la condition suivante :

$$(7) \quad \sum_{i=1}^r \lambda_i \Phi(x_i) = 0 \quad \forall \Phi \in \mathcal{D}_R$$

Dans le cas où $X=[a,b]$ intervalle fini de la droite réelle, nous allons démontrer, à l'aide des théorèmes du paragraphe précédent, un résultat qui généralise celui donné par BROSOWSKI [1965 b] et LOEB [1966] dans le cas de l'approximation rationnelle sans contraintes.

Théorème 12: toute fonction $f \in \tilde{C}[a,b]$ a dans $W_{m,n}$ un élément optimal unique si et seulement si $\mathcal{D}_R$ est un sous-espace de Haar pour tout $R \in W_{m,n}$.

(a) Supposons qu'il existe $f \in \tilde{C}[a,b]$ possédant dans $W_{m,n}$ deux éléments optimaux distincts.

$$R(x) = \frac{\sum_0^m a_i g_i(x)}{\sum_0^n b_i h_i(x)} \quad \text{et} \quad R^*(x) = \frac{\sum_0^m a_i^* g_i(x)}{\sum_0^n b_i^* h_i(x)}$$

Alors, d'après le théorème 2, un autre élément optimal pour f est :

$$\bar{R}(x) = \frac{\sum [\lambda a_i + (1-\lambda) a_i^*] g_i(x)}{\sum [\lambda b_i + (1-\lambda) b_i^*] h_i(x)} \quad \text{avec} \quad 0 < \lambda < 1$$

En outre, on a (voir théorème I-2) :

$$R(x) = \bar{R}(x) \quad \forall x \in E_{\bar{R}}$$

Considérons dans $\mathcal{D}_{\bar{R}}$ l'élément suivant :

$$\begin{aligned} \Phi(x) &= \sum_0^m a_i g_i(x) - \bar{R}(x) \sum_0^n b_i h_i(x) \\ &= [R(x) - \bar{R}(x)] \sum_0^n b_i h_i(x) \end{aligned}$$

$\Phi(x)$ n'est pas identiquement nul et s'annule pour tout $x \in E_{\bar{R}}$. D'après le théorème 10, $\Phi(x)$ a donc au moins $1+\eta(\mathcal{D}_{\bar{R}})$ zéros dans $[a,b]$, ce qui prouve que $\mathcal{D}_{\bar{R}}$ n'est pas un sous-espace de Haar.

(b) Réciproquement supposons qu'il existe un élément R de $W_{m,n}$ tel que $\mathcal{D}_R$ ne soit pas un sous-espace de Haar. Désignons par N la dimension

de $\mathcal{D}_R$ et soit $\Phi_1(x), \dots, \Phi_N(x)$ une base de $\mathcal{D}_R$. Il existe dans $\mathcal{D}_R$ un élément $\Phi(x)$ non identiquement nul qui a N zéros distincts $x_1, \dots, x_N$ dans $[a, b]$. Posons :

$$\Phi(x) \equiv \sum_1^N C_i^* \Phi_i(x) \equiv \sum_0^m a_i^* g_i(x) - R(x) \sum_0^n b_i^* h_i(x)$$

Le système linéaire homogène suivant a une solution non identiquement nulle :

$$(8) \quad \sum_{i=1}^N \tau_i \Phi_i(x_j) = 0 \quad j=1, \dots, N$$

Le déterminant de ce système est donc nul, ce qui prouve qu'il existe N nombres $\beta_1, \dots, \beta_N$ non tous nuls tels que :

$$\sum_{i=1}^N \beta_i \Phi_j(x_i) = 0 \quad j=1, \dots, N$$

En posant : $\beta_i = \theta_i \alpha_i$ avec $\theta_i \geq 0$, $|\alpha_i| = 1$

ceci prouve qu'il existe N nombres $\alpha_1, \dots, \alpha_N$ de module 1 tels que l'origine de R^N appartienne à l'enveloppe convexe de l'ensemble des points :

$$(9) \quad \{\alpha_i [\Phi_1(x_i), \dots, \Phi_N(x_i)] \mid i=1, \dots, N\}$$

Posons alors :

$$R(x) = \frac{\sum a_i g_i(x)}{\sum b_i h_i(x)} \quad \text{et} \quad \bar{R}(x) = \frac{\sum [\lambda a_i^* + (1-\lambda) a_i] g_i(x)}{\sum [\lambda b_i^* + (1-\lambda) b_i] h_i(x)}$$

avec : $0 < \lambda < 1$

Pour λ assez petit, le dénominateur de $\bar{R}(x)$ reste positif sur $[a, b]$ et donc $\bar{R} \in W_{m,n}$.

D'autre part :

$$R(x) - \bar{R}(x) = \frac{-\lambda \Phi(x)}{\sum [\lambda b_i^* + (1-\lambda) b_i] h_i(x)}$$

On a donc :

$$R(x_i) = \bar{R}(x_i) \quad i = 1, \dots, N$$

et, pour λ assez petit :

$$\|R - \bar{R}\| \leq 1.$$

Considérons alors la fonction :

$$f(x) = R(x) + [1 - |R(x) - \bar{R}(x)|] g(x)$$

avec :

$$g \in C[a, b] \quad , \quad \|g\| = 1 \quad , \quad g(x_i) = \alpha_i$$

On a alors (comme dans la démonstration du théorème II-6) :

$$\begin{cases} f(x_i) - R(x_i) = f(x_i) - \bar{R}(x_i) = \alpha_i \\ \|f - R\| = \|f - \bar{R}\| = 1 \end{cases}$$

La formule (9) prouve donc, d'après le théorème 5, que R et $\bar{R}$ sont deux éléments optimaux distincts pour f , ce qui achève la démonstration.

Remarque 1 : il y a équivalence entre le théorème 12 et le théorème 11 (dans lequel on prend $X = [a, b]$) en vertu du lemme suivant :

Lemme : $R \in W_{m,n}$ possède la propriété (P) du théorème 11 si et seulement si $\mathcal{D}_R$ est un sous-espace de Haar de $C[a, b]$.

Posons $N = \dim(\mathcal{D}_R)$ et soit $\Phi_1(x), \dots, \Phi_N(x)$ une base de $\mathcal{D}_R$.

(a) Supposons que $\mathcal{D}_R$ soit un sous-espace de Haar et que R ne possède pas la propriété (P). Alors il existe des points x_i de $[a, b]$ et des nombres λ_i tels que :

$$(10) \quad \begin{cases} \lambda_i \neq 0 \quad , \quad 1 \leq i \leq N \\ \sum_{i=1}^r \lambda_i \Phi(x_i) = 0 \quad \forall \Phi \in \mathcal{D}_R \end{cases}$$

ou encore :

$$(11) \quad \sum_{i=1}^r \lambda_i \Phi_j(x_i) = 0 \quad j = 1, \dots, N$$

Si $r < N$, complétons l'ensemble $\{x_1, \dots, x_r\}$ en un ensemble $\{x_1, \dots, x_N\}$ de N points de $[a, b]$ deux à deux distincts et considérons le système linéaire suivant :

$$\sum_{i=1}^N \tau_i \Phi_j(x_i) = 0 \quad j = 1, \dots, N$$

D'après (11), ce système admet une solution non identiquement nulle et par conséquent son déterminant est nul :

$$(12) \quad \begin{vmatrix} \Phi_1(x_1) & \dots & \Phi_1(x_N) \\ \dots & \dots & \dots \\ \Phi_N(x_1) & \dots & \Phi_N(x_N) \end{vmatrix} = 0$$

ce qui est impossible puisque par hypothèse $\mathcal{D}_R$ est un sous-espace de Haar.

Par conséquent, si $\mathcal{D}_R$ est un sous-espace de Haar de $C[a, b]$, $R(x)$ possède la propriété (P).

(b) Réciproquement, supposons que R possède la propriété (P) et que $\mathcal{D}_R$ ne soit pas un sous-espace de Haar de $C[a, b]$.

Il existe alors dans $\mathcal{D}_R$ un élément $\Phi(x)$ non identiquement nul possédant dans $[a, b]$ N zéros distincts $x_1, \dots, x_N$.

On a alors la relation (12) qu'on peut écrire :

$$\sum_{i=1}^N \lambda_i \begin{bmatrix} \Phi_1(x_i) \\ \vdots \\ \Phi_N(x_i) \end{bmatrix} = 0$$

avec des coefficients λ_i non tous nuls. D'où l'on déduit :

$$\sum_{i=1}^N \lambda_i \Phi(x_i) = 0 \quad \forall \Phi \in \mathcal{D}_R$$

ou encore :

$$\sum_{i=1}^r \lambda_i \Phi(x_i) = 0 \quad \forall \Phi \in \mathcal{D}_R$$

avec : $1 \leq r \leq N$ et $\lambda_i \neq 0$.

D'après (7), ceci montre que R ne possède pas la propriété (P) ce qui est contraire à l'hypothèse.

Par conséquent, si $R(x)$ possède la propriété (P), $\mathcal{D}_R$ est un sous-espace de Haar, ce qui termine la démonstration du lemme.

Remarque 2: En tenant compte du lemme ci-dessus, le théorème II-7 nous donne la condition suffisante d'unicité ci-dessous:

Théorème 13: soit $R \in W_{m,n}$ optimal pour $f \in C[X]$. Si R possède la propriété (P), (ou, lorsque $X=[a,b]$, si $\mathcal{D}_R$ est un sous-espace de Haar), R est le seul élément optimal pour f dans $W_{m,n}$.

Remarque 3: lorsque $X=[a,b]$ et $\mathcal{D}_R$ est un sous-espace de Haar quel que soit $R \in W_{m,n}$, les théorèmes 9, 10 et 12 montrent que toute fonction $f \in \tilde{C}[a,b]$ possède dans $W_{m,n}$ un élément optimal unique caractérisé par le fait que $f-R$ possède au moins $[1 + \dim(\mathcal{D}_R)]$ (e) - points alternés.

Ce résultat peut aussi s'obtenir en partant des résultats de MEINARDUS ([1964], théorèmes 86 et 87).

6- Unicité forte.

Dans ce paragraphe, nous supposerons encore que $X=[a,b]$ et en outre nous prendrons les contraintes (1) homogènes c'est-à-dire:

$$c_i = 0 \quad i = 1, \dots, K$$

Soit $R^* \in W_{m,n}$ optimal pour f . D'après le théorème 13, si $\nu(\mathcal{D}_{R^*}) = \eta(\mathcal{D}_{R^*})$, on a:

$$\|f-R\| - \|f-R^*\| > 0 \quad \forall R \in W_{m,n}, \quad R \neq R^*$$

Nous allons montrer que si en outre $\dim(\mathcal{D}_{R^*}) = m+n+1-K$, alors:

$$\|f-R\| - \|f-R^*\| \geq \gamma \|R-R^*\|$$

où γ est une constante strictement positive.

Nous allons tout d'abord démontrer un lemme préliminaire:

Lemme: soit $R^* = P^*/Q^*$ un élément de $W_{m,n}$ tel que $\dim(\mathcal{D}_{R^*}) = m+n+1-K$ et soit $\Phi(x) = P(x) - R^*(x) Q(x)$ un élément de $\mathcal{D}_{R^*}$ tel que :

$$\Phi(x) \equiv 0, \quad \|P^*\| + \|Q^*\| = \|P\| + \|Q\| \quad \text{et} \quad Q(x) \geq 0 \quad \text{sur} \quad [a, b].$$

Alors : $P(x) \equiv P^*(x) \quad \text{et} \quad Q(x) \equiv Q^*(x).$

Si $R^* \equiv 0$, on a alors : $P \equiv P^* \equiv 0$

D'où l'on déduit :

$$\|Q\| = \|Q^*\|$$

avec :

$$\begin{cases} n = 0 \\ Q(x) Q^*(x) \geq 0 \quad \text{sur} \quad [a, b] \end{cases}$$

ce qui prouve que : $Q(x) \equiv Q^*(x).$

Dans le cas où $R^* \neq 0$, les relations :

$$\sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = 0 \quad (i=1, \dots, K)$$

permettent de calculer K des coefficients a_j, b_j en fonction des autres :

$$(13) \quad \begin{cases} a_j = \sum_{i=r}^m \gamma_{ij} a_i + \sum_{i=K-r}^n \delta_{ij} b_i & j=0, \dots, r-1 \\ b_j = \sum_{i=r}^m \gamma'_{ij} a_i + \sum_{i=K-r}^n \delta'_{ij} b_i & j=0, \dots, K-r-1 \end{cases}$$

Tout élément de $\mathcal{D}_{R^*}$ peut alors s'écrire sous la forme :

$$\Phi(x) = a_r \Phi_1(x) + \dots + a_m \Phi_{m-r+1}(x) + b_{K-r} \Phi_{m-r+2}(x) + \dots + b_r \Phi_{m+n+2-K}(x)$$

Mais on a :

$$\Phi^*(x) \equiv P^*(x) - R^*(x) Q^*(x) \equiv 0$$

c'est-à-dire :

$$a_r^* \Phi_1(x) + \dots + b_n^* \Phi_{m+n+2-K}(x) \equiv 0$$

L'un des coefficients $a_r^*, \dots, a_m^*, b_{K-r}^*, \dots, b_n^*$ est non nul (sinon d'après (13), tous les a_i^*, b_i^* seraient nuls). Supposons donc $b_n^* \neq 0$. Tout élément $\Phi(x)$ de $\mathcal{D}_{R^*}$ s'écrit alors :

$$\begin{aligned} \Phi(x) = & \left(a_r - \frac{b_n}{b_n^*} a_r^* \right) \Phi_1(x) + \dots + \left(a_m - \frac{b_n}{b_n^*} a_m^* \right) \Phi_{m-r+1}(x) \\ & + \left(b_{K-r} - \frac{b_n}{b_n^*} b_{K-r}^* \right) \Phi_{m-r+2}(x) + \dots + \left(b_{n-1} - \frac{b_n}{b_n^*} b_{n-1}^* \right) \Phi_{m+n+1-K}(x) \end{aligned}$$

Compte tenu de la relation: $\dim(\mathcal{D}_{R^*}) = m+n+1-K$, cela prouve que $\Phi_1(x), \dots, \Phi_{m+n+1-K}(x)$ est une base de $\mathcal{D}_{R^*}$.

Or par hypothèse, on a:

$$P(x) - R^*(x) Q(x) \equiv 0$$

Donc les coefficients de P et Q vérifient:

$$\begin{aligned} a_j - \frac{b_n}{b_n^*} a_j^* &= 0 & j=r, \dots, m \\ b_j - \frac{b_n}{b_n^*} b_j^* &= 0 & j=K-r, \dots, n-1 \end{aligned}$$

D'où:

$$\begin{cases} a_j = \lambda a_j^* & j=0, \dots, m \\ b_j = \lambda b_j^* & j=0, \dots, n \end{cases}$$

On a donc:

$$P(x) \equiv \lambda P^*(x) \quad , \quad Q(x) \equiv \lambda Q^*(x)$$

En tenant compte des relations:

$$\|P\| + \|Q\| = \|P^*\| + \|Q^*\|$$

$$Q(x) Q^*(x) \geq 0 \quad \text{sur } [a, b]$$

on voit que $\lambda = 1$, ce qui termine la démonstration du lemme.

Nous allons maintenant démontrer le théorème suivant qui généralise celui donné par CHENEY [1966] dans le cas où il n'y a pas de contraintes sur les coefficients.

Théorème 14: soit R^* optimal pour f dans $W_{m,n}$. Si $\eta(\mathcal{D}_{R^*}) = m+n+1-K$, il existe une constante $\gamma > 0$ telle que:

$$\|f - R\| \geq \|f - R^*\| + \gamma \|R - R^*\| \quad \forall R \in W_{m,n}$$

Si $f \in W_{m,n}$, on a $R^* \equiv f$ et on peut prendre $\gamma = 1$.

Supposons que $f \notin W_{m,n}$ et pour $R \in W_{m,n}$, $R \neq R^*$, posons :

$$\gamma(R) = \frac{\|f-R\| - \|f-R^*\|}{\|R-R^*\|}$$

Il faut montrer que $\gamma(R)$ reste borné inférieurement par un nombre strictement positif. Supposons au contraire qu'il existe une suite $R_1(x), \dots, R_j(x), \dots$ d'éléments de $W_{m,n}$ tels que :

$$R_j \neq R^* \quad \text{et} \quad \gamma(R_j) \rightarrow 0$$

Posons :

$$R_j(x) = \frac{P_j(x)}{Q_j(x)}, \quad R^*(x) = \frac{P^*(x)}{Q^*(x)}$$

Comme les contraintes sont homogènes, on peut en outre supposer :

$$\|P_j\| + \|Q_j\| = \|P^*\| + \|Q^*\| = 1$$

Il en résulte que les coefficients de $P_j(x)$ et $Q_j(x)$ sont bornés et donc, de la suite initiale, on peut extraire une sous-suite telle que :

$$P_j(x) \rightarrow P(x), \quad Q_j(x) \rightarrow Q(x)$$

Puisque $\gamma(R_j) \rightarrow 0$, $\|R_j\|$ est borné et donc :

$$\gamma(R_j) \|R_j - R^*\| \rightarrow 0 \quad \text{quand} \quad j \rightarrow \infty$$

Posons :

$$h(x) = \text{signe} [f(x) - R^*(x)]$$

Considérons un point $x \in E_{R^*}$:

$$\begin{aligned} \gamma(R_j) \|R_j - R^*\| &= \|f - R_j\| - \|f - R^*\| \\ &= \|f - R_j\| - h(x) [f(x) - R^*(x)] \end{aligned}$$

D'où :

$$\gamma(R_j) \|R_j - R^*\| \geq h(x) [f(x) - R_j(x)] - h(x) [f(x) - R^*(x)]$$

$$(14) \quad \gamma(R_j) \|R_j - R^*\| \geq - \frac{h(x)}{Q_j(x)} [P_j(x) - R^*(x) Q_j(x)]$$

Passons à la limite pour $j \rightarrow \infty$:

$$(15) \quad h(x) [P(x) - R^*(x) Q(x)] \geq 0 \quad \forall x \in E_{R^*}$$

Mais on a :

$$m+n+1-K = \eta(\mathcal{D}_{R^*}) \leq \dim(\mathcal{D}_{R^*}) \leq m+n+1-K$$

ce qui prouve que $\mathcal{D}_{R^*}$ est un sous-espace de Haar.

D'après le théorème 10, $\mathcal{J}-R^*$ possède dans $[a,b]$ au moins $N+1 = m+n+2-K$ (e)-points alternés. Notons les :

$$x_1 < \dots < x_{N+1}$$

et posons :

$$\Phi(x) = P(x) - R^*(x) \quad Q(x)$$

On a donc :

$$(-1)^i \varepsilon \Phi(x_i) \geq 0 \quad i = 1, \dots, N+1 \quad (\varepsilon = \pm 1) .$$

Supposons que :

$$\begin{aligned} \Phi(x_i) &\neq 0, & \Phi(x_{i+j+1}) &\neq 0 & (j \geq 0) \\ \Phi(x_r) &= 0 & i < r < i+j+1 \end{aligned}$$

On a :

$$(-1)^{j+1} \Phi(x_i) \Phi(x_{i+j+1}) > 0$$

Ceci prouve, en comptant deux fois les zéros intérieurs à $[a,b]$ où Φ ne change pas de signe, que $\Phi(x)$ a au moins $(j+1)$ zéros dans $]x_i, x_{i+j+1}[$.

Au total, $\Phi(x)$ a donc au moins N zéros dans $[a,b]$.

Comme $\mathcal{D}_{R^*}$ est un sous-espace de Haar de dimension N , ceci prouve (d'après le théorème 0-3) que :

$$P(x) - R^*(x) \quad Q(x) \equiv 0 .$$

En appliquant le lemme précédent, on obtient donc :

$$P(x) \equiv P^*(x) \quad , \quad Q(x) \equiv Q^*(x)$$

Puisque $Q^*(x) > 0$ sur $[a,b]$, on peut supposer que l'on a pour tout j et pour un certain $\varepsilon > 0$:

$$Q_j(x) \geq \varepsilon \quad \text{sur } [a,b]$$

Posons :

$$\begin{aligned} d &= \inf_{\Phi \in \mathcal{D}_{R^*}} \max_{x \in E_{R^*}} [h(x) \quad \Phi(x)] \\ &\|\Phi\| = 1 \end{aligned}$$

Pour $\|\Phi\| = 1$, on a :

$$\max_{E_{R^*}} [h(x) \quad \Phi(x)] > 0$$

Sinon, on aurait :

$$h(x) \Phi(x) \leq 0 \quad \forall x \in E_{R^*}$$

et comme ci-dessus on montrerait que $\Phi(x) \equiv 0$, en contradiction avec $\|\Phi\|=1$.

On en déduit que $d > 0$. Car si d était nul, on pourrait trouver une suite d'éléments Φ_j de $\mathcal{D}_{R^*}$ convergente vers un élément Φ tel que :

$$h(x) \Phi(x) \leq 0 \quad \forall x \in E_{R^*}$$

avec $\|\Phi\| = 1$, ce qui est impossible.

D'après (14), on a :

$$\forall (R_j) \quad \|R_j - R^*\| \geq \frac{h(x)}{Q_j(x)} [Q_j(x) R^*(x) - P_j(x)] \quad \forall x \in E_{R^*}$$

D'après ce qui précède, il y a au moins un point $x \in E_{R^*}$ tel que le second membre de l'inégalité ci-dessus soit strictement positif. Prenons nous en ce point :

$$\begin{aligned} \forall (R_j) \quad \|R_j - R^*\| &\geq h(x) [Q_j(x) R^*(x) - P_j(x)] \\ &\geq d \| (R^* - R_j) \cdot Q_j \| \\ &\geq d \varepsilon \|R^* - R_j\| \quad \text{car } \varepsilon \leq Q_j(x) \leq 1 \end{aligned}$$

Comme $R_j \neq R^*$, on en déduit :

$$\forall (R_j) \quad \geq d \varepsilon$$

en contradiction avec l'hypothèse $\forall (R_j) \rightarrow 0$, ce qui démontre le théorème.

7- Continuité.

Dans ce paragraphe également, nous supposons $X = [a, b]$ et nous prenons des contraintes homogènes :

$$\sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = 0 \quad i = 1, \dots, K$$

Etant donné $f \in C[a, b]$, notons $\mathcal{T}_f$ l'ensemble des fractions de $W_{m,n}$ qui réalisent la meilleure approximation de f (cet ensemble $\mathcal{T}_f$ peut éventuellement être vide).

Nous allons démontrer un résultat qui généralise celui donné dans le cas sans contraintes par CHENEY ([1964], [1966]).

Théorème 15: si $R \in \mathcal{T}f_0$ et $\eta(\mathcal{D}_{R_0}) = m+n+1-K$, alors $\mathcal{T}f$ est non vide pour tous les f d'un voisinage de f_0 et $\mathcal{T}$ est "continu" en f_0 c'est-à-dire qu'il existe $\beta > 0$ tel que :

$$\|R - R_0\| < \beta \|f - f_0\| \quad \forall R \in \mathcal{T}f$$

Soit $f \in C[a, b]$. Pour chercher un élément optimal pour f , on peut se borner aux éléments $R \in W_{m,n}$ tels que :

$$\|f - R\| \leq \|f - R_0\|$$

D'après le théorème 14, on a :

$$\begin{aligned} \forall \|R - R_0\| &\leq \|f_0 - R\| - \|f_0 - R_0\| \\ &\leq \|f_0 - f\| + \|f - R\| - \|f_0 - R_0\| \\ &\leq \|f_0 - f\| + \|f - R_0\| - \|f_0 - R_0\| \\ &\leq 2 \|f - f_0\| \end{aligned}$$

Donc s'il existe un élément $R \in \mathcal{T}f$, cet élément vérifie :

$$\|R - R_0\| \leq \frac{2}{\gamma} \|f - f_0\|$$

ce qui prouve l'existence de β .

Montrons maintenant que pour $\|f - f_0\|$ assez petit, $\mathcal{T}f$ n'est pas vide.

Posons :

$$R_0(x) = \frac{P_0(x)}{Q_0(x)}$$

avec :

$$\|P_0\| + \|Q_0\| = 1$$

et :

$$\inf_{[a,b]} Q_0(x) = 2\varepsilon_1 > 0$$

Il existe un nombre $\varepsilon_2 > 0$ tel que :

$$\left. \begin{aligned} R = \frac{P}{Q} \in W_{m,n} \\ \|P\| + \|Q\| = 1 \\ \|R - R_0\| < \varepsilon_2 \end{aligned} \right\} \implies \|Q - Q_0\| < \varepsilon_1$$

Sinon on pourrait trouver dans $W_{m,n}$ une suite d'éléments $R_j = P_j / Q_j$ tels que :

$$\begin{aligned} R_j &\rightarrow R_0 \\ \|P_j\| + \|Q_j\| &= 1 \\ \|Q_j - Q_0\| &\geq \varepsilon_1 \end{aligned}$$

De cette suite, on pourrait extraire une sous-suite convergente :

$$P_j \rightarrow P, \quad Q_j \rightarrow Q$$

On aurait alors :

$$\begin{aligned} P(x) - R_0(x) Q(x) &\equiv 0 \\ \|P\| + \|Q\| &= \|P_0\| + \|Q_0\| = 1 \\ Q(x) &\geq 0 \quad \text{sur } [a, b] \end{aligned}$$

d'où d'après le lemme du paragraphe précédent :

$$P \equiv P_0, \quad Q \equiv Q_0$$

en contradiction avec l'inégalité :

$$\|Q - Q_0\| \geq \varepsilon_1.$$

L'existence de ε_2 étant vérifiée, montrons maintenant que $\mathcal{C}f$ est non vide lorsque :

$$\|f - f_0\| < \frac{1}{2} \varepsilon_2$$

S'il existe dans $W_{m,n}$ un élément R optimal pour f , on a tout d'abord :

$$\|R - R_0\| \leq \frac{2}{\gamma} \|f - f_0\| < \varepsilon_2$$

On peut en outre imposer à R la condition :

$$\|P\| + \|Q\| = 1$$

On aura alors, d'après ce qui précède :

$$\|Q - Q_0\| < \varepsilon_1$$

et donc, d'après la définition de ε_1 :

$$Q(x) \geq \varepsilon_1 \quad \text{sur } [a, b]$$

Donc pour chercher un élément optimal pour f , on peut se borner au sous-ensemble suivant de $W_{m,n}$:

$$\left\{ R = \frac{P}{Q} \mid \|P\| + \|Q\| = 1, \quad Q(x) \geq \varepsilon_1 \quad \text{sur } [a, b] \right\}$$

Il lui correspond dans l'espace des coefficients un ensemble compact et donc il existe au moins un élément optimal pour f . Ce qui termine la démonstration du théorème.

8- Retour sur le cas des contraintes inégalités.

Revenons au problème étudié dans le chapitre précédent, c'est-à-dire l'approximation d'une fonction $f \in C[a, b]$ par la classe $W_{m, n}$ des fractions rationnelles

$$R(x) = \frac{\sum_0^m a_j g_j(x)}{\sum_0^n b_j h_j(x)}$$

dont les coefficients a_j, b_j vérifient les contraintes suivantes :

$$\left\{ \begin{array}{l} \sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i \quad i = 1, \dots, K \\ \sum_{j=0}^m a_j \alpha_j(t) + \sum_{j=0}^n b_j \beta_j(t) \geq c(t) \quad \forall t \in G \\ \sum_{j=0}^n b_j h_j(x) > 0 \quad \text{sur } [a, b] \end{array} \right.$$

Soit R optimal pour f dans $W_{m, n}$. Les théorèmes II-5 (remarque 3) et III-6 montrent qu'il existe des points $t_1, \dots, t_s$ ($s \geq 0$) de G tels que $R(x)$ soit optimal pour f dans la classe $V(t_1, \dots, t_s)$ des fractions rationnelles

$$R(x) = \frac{\sum_0^m a_j g_j(x)}{\sum_0^n b_j h_j(x)}$$

dont les coefficients a_j, b_j vérifient les conditions :

$$(16) \quad \sum_{j=0}^m \alpha_{ij} a_j + \sum_{j=0}^n \beta_{ij} b_j = c_i \quad i = 1, \dots, K$$

$$(17) \quad \sum_{j=0}^m a_j \alpha_j(t_i) + \sum_{j=0}^n b_j \beta_j(t_i) = c(t_i) \quad i = 1, \dots, s$$

$$(18) \quad \sum_{j=0}^n b_j h_j(x) > 0 \quad \text{sur} \quad [a, b]$$

Désignons par $F(R; t_1, \dots, t_s)$ la classe des fonctions

$$\Phi(x) = \sum_{j=0}^m a_j g_j(x) - R(x) \sum_{j=0}^n b_j h_j(x)$$

dont les coefficients a_j, b_j vérifient (16) et (17).

Aux résultats (A) et (B) du chapitre précédent (paragraphe 5), le théorème 10 nous permet d'ajouter le résultat :

(C) soit $R(x) = \sum a_j^* g_j(x) / \sum b_j^* h_j(x)$ optimal pour f dans $W_{m,n}$:

1) Il existe s points $t_1, \dots, t_s$ ($s \geq 0$) de G tels que

$$\sum_{j=0}^m a_j^* \alpha_j(t_i) + \sum_{j=0}^n b_j^* \beta_j(t_i) = c(t_i) \quad i = 1, \dots, s$$

2) Le nombre de (e) -points de signes alternés de $f - R$ est au moins égal à $1 + \eta[F(R; t_1, \dots, t_s)]$.

Enfin le théorème 13 va nous permettre de donner une autre forme du théorème II-7 :

(D) soit $R \in W_{m,n}$ optimal pour f . Si $F(R; t_1, \dots, t_s)$ est un sous-espace de Haar de $C[a, b]$, alors R est le seul élément optimal pour f dans $W_{m,n}$.

En effet, R étant optimal, il existe des points $t_1, \dots, t_s$ de G , des (e) -points $x_1, \dots, x_r$ ($r \geq 1$) de $f - R$ et des nombres λ_i, μ_i tels que :

$$(19) \quad \lambda_i [f(x_i) - R(x_i)] > 0, \quad \mu_i > 0$$

$$(20) \quad \sum_{i=1}^r \lambda_i \Phi(x_i) + \sum_{i=1}^s \mu_i \left[\sum_{j=0}^m a_j \alpha_j(t_i) + \sum_{j=0}^n b_j \beta_j(t_i) - c(t_i) \right] = 0$$

pour tout $\Phi \in \Delta_R$

$F(R; t_1, \dots, t_s)$ étant un sous-espace de Haar, d'après le théorème 13, R est le seul élément optimal pour f dans la classe $V(t_1, \dots, t_s)$ définie au début de ce paragraphe.

Supposons que f possède dans $W_{m,n}$ un élément optimal $R' \neq R$.
Posons :

$$R'(x) = \frac{\sum_0^m a_j' g_j(x)}{\sum_0^n b_j' h_j(x)}$$

et considérons dans Δ_R l'élément suivant :

$$\Phi(x) = \sum_0^m a_j' g_j(x) - R(x) \sum_0^n b_j' h_j(x)$$

On a :

$$\Phi(x) = \{[f(x) - R(x)] - [f(x) - R'(x)]\} \sum_0^n b_j' h_j(x)$$

et, quel que soit $x \in E_R$:

$$|f(x) - R'(x)| \leq \|f - R'\| = \|f - R\| = |f(x) - R(x)|$$

$$\text{D'où :} \quad \Phi(x)[f(x) - R(x)] \geq 0 \quad \forall x \in E_R$$

Et donc d'après (19) :

$$\lambda_i \Phi(x_i) \geq 0 \quad i = 1, \dots, r$$

De plus, comme $R' \in W_{m,n}$, on a :

$$\mu_i \left[\sum_{j=0}^m a_j' \alpha_j(t_i) + \sum_{j=0}^n b_j' \beta_j(t_i) - c(t_i) \right] \geq 0 \quad i = 1, \dots, s$$

D'où l'on déduit en reportant dans (20) :

$$\Phi(x_i) = 0 \quad i = 1, \dots, r$$

$$(21) \quad \sum_{j=0}^m a_j' \alpha_j(t_i) + \sum_{j=0}^n b_j' \beta_j(t_i) = c(t_i) \quad i = 1, \dots, s$$

La relation (21) prouve que $R'(x)$ est un élément de la classe $V(t_1, \dots, t_s)$. Il y a contradiction puisque, dans cette classe, f n'admet que R comme élément optimal.

Ce qui prouve que R' ne peut pas exister et donc que R est le seul élément optimal pour f dans $W_{m,n}$.

APPROXIMATION RATIONNELLE AVEC CONDITIONS DE NOEUDS

Dans ce chapitre, on étudie l'approximation d'une fonction continue $f(x)$ par des fractions rationnelles généralisées :

$$R(x) = \frac{\sum_0^m a_j g_j(x)}{\sum_0^n b_j h_j(x)}$$

prenant des valeurs données en des points donnés $\xi_1, \dots, \xi_K$.

Le problème est posé au paragraphe 1. Dans les paragraphes 2 et 3, on examine ce que deviennent les théorèmes sur l'existence et la caractérisation des éléments optimaux. Au paragraphe 4 sont étudiées les propriétés d'alternance. Dans les paragraphes 5 et 6, on examine l'unicité et "l'unicité forte", la continuité étant étudiée au paragraphe 7. Dans le paragraphe 8, on examine certains cas particuliers dans lesquels on peut appliquer les résultats du présent chapitre et aussi ceux du chapitre précédent. Au paragraphe 9, on complète dans certains cas particuliers les résultats énoncés au chapitre II.

1 - Introduction.

On considère ici encore un espace compact X et $C[X]$ espace des fonctions continues sur X , à valeurs réelles. On se donne dans $C[X]$ des fonctions $g_0(x), \dots, g_m(x), h_0(x), \dots, h_n(x)$ et on suppose les $g_i(x)$ d'une part, les $h_i(x)$ d'autre part linéairement indépendantes sur X .

On se donne en outre un compact X' contenant X et, dans X' , un ensemble $T = \{\xi_1, \dots, \xi_K\}$ de points distincts appelés *noeuds*.

On désigne par $W_{m,n}$ la classe des fonctions :

$$R(x) = \frac{\sum_0^m a_j g_j(x)}{\sum_0^n b_j h_j(x)}$$

où les a_j, b_j sont des nombres réels tels que :

$$(1) \quad R(\xi_i) = \theta_i \quad i = 1, \dots, K$$

$$(2) \quad \sum_{j=0}^n b_j h_j(x) > 0 \quad \text{sur } X$$

$$(3) \quad \sum_{j=0}^n b_j h_j(\xi_i) \neq 0 \quad i = 1, \dots, K$$

$\theta_1, \dots, \theta_K$ sont des nombres réels donnés et on suppose que les vecteurs V_i ($i = 1, \dots, K$) de composantes

$$g_0(\xi_i), \dots, g_m(\xi_i), -\theta_i h_0(\xi_i), \dots, -\theta_i h_n(\xi_i)$$

sont linéairement indépendants.

Etant donné $f \in C[X]$, on cherche dans $W_{m,n}$ un élément R^* tel que :

$$\|f - R^*\| \leq \|f - R\| \quad \forall R \in W_{m,n}$$

Comme précédemment, nous utiliserons les notations suivantes :

$$y_R(x) = f(x) - R(x)$$

$$\rho = \inf_{R \in W_{m,n}} \|f - R\|$$

$$E_R = \{x \in X \mid f(x) - R(x) = \pm \|f - R\|\}$$

Ici encore, les classes Δ_R et $\mathcal{D}_R$ définies au chapitre II sont confondues. C'est l'ensemble des fonctions :

$$\Phi(x) = \sum_0^m a_j' g_j(x) - R(x) \sum_0^n b_j' h_j(x)$$

dont les coefficients a_j', b_j' vérifient les relations :

$$(4) \quad \sum_{j=0}^m a_j' g_j(\xi_i) - \theta_i \sum_{j=0}^n b_j' h_j(\xi_i) = 0 \quad i = 1, \dots, K$$

qui s'écrivent simplement, en tenant compte de (1) :

$$(5) \quad \Phi(\xi_i) = 0 \quad i = 1, \dots, K$$

2- Existence d'un élément optimal.

Comme précédemment, nous désignons par $\overline{W}_{m,n}$ la classe des fonctions $R \in C[X]$ qui vérifient une identité de la forme :

$$R(x) \sum_{j=0}^n b_j h_j(x) = \sum_{j=0}^m a_j g_j(x)$$

où les a_j, b_j sont des nombres réels non tous nuls vérifiant les relations (4).

Les théorèmes 1, 2 et 3 du chapitre précédent nous donnent tout d'abord les résultats suivants :

Théorème 1 : si les fonctions $h_0(x), h_1(x), \dots, h_n(x)$ sont linéairement indépendantes sur tout sous-ensemble ouvert de X , alors toute fonction $f \in C[X]$ a dans $\overline{W}_{m,n}$ au moins un élément optimal.

Théorème 2 : l'ensemble des points de R^{m+n+2} dont les coordonnées $(a_0, a_1, \dots, a_m, b_0, b_1, \dots, b_n)$ sont les coefficients des éléments $R \in W_{m,n}$ optimaux pour f est convexe.

Théorème 3 : soit $R \in W_{m,n}$. S'il existe un ensemble D de points de X tel que :

$$a) \quad f(x) - R(x) \neq 0 \quad \forall x \in D$$

$$b) \quad \inf_{x \in D} \Phi(x) [f(x) - R(x)] \leq 0 \quad \forall \Phi \in \mathcal{D}_R$$

alors, on a :

$$\rho \geq \inf_{x \in D} |f(x) - R(x)|$$

Ce théorème est aussi valable sous la forme du corollaire du théorème I-3 c'est-à-dire :

S'il existe des points $x_1, \dots, x_r$ de X et des nombres $\lambda_1, \dots, \lambda_r$ tels que :

$$\lambda_i [f(x_i) - R(x_i)] > 0 \quad i = 1, \dots, r$$

$$\sum_{i=1}^r \lambda_i \Phi(x_i) \leq 0 \quad \forall \Phi \in \mathcal{D}_R$$

alors on a :

$$\rho \geq \min_{1 \leq i \leq r} |f(x_i) - R(x_i)|$$

En particulier, supposons qu'il existe des noeuds ξ_i appartenant à X .

$$\text{On a : } \Phi(\xi_i) = 0 \quad \forall \Phi \in \mathcal{D}_R$$

et en appliquant le résultat ci-dessus, on retrouve l'inégalité évidente :

$$\rho \geq |f(\xi_i) - \theta_i|$$

D'où :

$$(6) \quad \rho \geq \max_{\xi_i \in X} |f(\xi_i) - \theta_i|$$

3- Caractérisation des éléments optimaux.

Les théorèmes de caractérisation se déduisent immédiatement des théorèmes 4 à 7 du chapitre précédent.

Remarquons que si E_R contient un noeud, c'est-à-dire si :

$$\|f - R\| = \max_{\xi_i \in X} |f(\xi_i) - \theta_i|$$

alors R est optimal pour f dans $W_{m,n}$. Ceci nous permettra d'énoncer les théorèmes de caractérisation sous une forme légèrement modifiée :

Théorème 4 : pour que $R \in W_{m,n}$ soit optimal pour f , il faut et il suffit que l'une des deux conditions suivantes soit vérifiée :

(a) E_R contient un noeud c'est-à-dire :

$$\|f - R\| = \max_{\xi_i \in X} |f(\xi_i) - \theta_i|$$

(b) E_R ne contient aucun noeud et on a :

$$\inf_{x \in E_R} \Phi(x) [f(x) - R(x)] \leq 0 \quad \forall \Phi \in \mathcal{D}_R$$

D'après le théorème III-5, la condition (b) ci-dessus peut être remplacée par la condition :

(b') E_R ne contient aucun noeud et l'origine de R^N appartient à l'enveloppe convexe de l'ensemble de points :

$$\{[\psi_R(x) \ \Phi_1(x), \dots, \psi_R(x) \ \Phi_N(x)] \in R^N \mid x \in E_R\}$$

où $\Phi_1(x), \dots, \Phi_N(x)$ est une base arbitraire de $\mathcal{D}_R$.

Théorème 5 : pour que $R \in W_{m,n}$ soit optimal pour f , il faut et il suffit que l'une des deux conditions suivantes soit vérifiée :

$$(c) \quad \|f - R\| = \max_{\xi_i \in X} |f(\xi_i) - \theta_i|$$

(d) Il existe des (e)-points $x_1, \dots, x_r$ de $f - R$, distincts des noeuds, et des nombres $\lambda_1, \dots, \lambda_r$ tels que :

$$1 \leq r \leq 1 + \dim(\mathcal{D}_R)$$

$$\lambda_i [f(x_i) - R(x_i)] > 0 \quad i = 1, \dots, r$$

$$\sum_{i=1}^r \lambda_i \Phi(x_i) = 0 \quad \forall \Phi \in \mathcal{D}_R$$

La démonstration du théorème II-5 a montré que la condition (d) ci-dessus peut être remplacée par la condition :

(d') Il existe des (e)-points $x_1, \dots, x_r$ de $f - R$, distincts des noeuds, et des nombres λ_i, γ_i tels que :

$$1 \leq r \leq 1 + \dim(\mathcal{D}_R)$$

$$\lambda_i [f(x_i) - R(x_i)] > 0 \quad i = 1, \dots, r$$

$$\sum_{i=1}^r \lambda_i \begin{bmatrix} g_0(x_i) \\ \vdots \\ g_m(x_i) \\ -R(x_i)h_0(x_i) \\ \vdots \\ -R(x_i)h_n(x_i) \end{bmatrix} + \sum_{i=1}^K \gamma_i \begin{bmatrix} g_0(\xi_i) \\ \vdots \\ g_m(\xi_i) \\ -\theta_i h_0(\xi_i) \\ \vdots \\ -\theta_i h_n(\xi_i) \end{bmatrix} = 0$$

Enfin, le théorème III-7 subsiste sans aucune modification :

Théorème 6 : pour que $R \in W_{m,n}$ soit optimal pour f , il faut et il suffit qu'il existe une fonctionnelle linéaire L telle que :

$$L(f - R) = \|f - R\| \quad , \quad \|L\| = 1$$

$$L(\Phi) = 0 \quad \forall \Phi \in \mathcal{D} \quad (\text{voir p. 149})$$

4- Propriétés d'alternance.

Dans ce paragraphe, nous supposons : $X = [a, b]$ intervalle fini de la droite réelle.

Les propriétés d'alternance démontrées au chapitre précédent sont évidemment encore valables mais nous allons obtenir d'autres résultats qui sont (nous le verrons plus loin) équivalents à ceux du chapitre précédent lorsque tous les noeuds sont extérieurs à $[a, b]$ et meilleurs lorsque $[a, b]$ contient au moins un noeud.

Nous allons utiliser une suite de nombres $\tau_1, \dots, \tau_N$ que nous définirons d'une manière analogue à celle employée par PASZKOWSKI [1955] dans le cas de l'approximation polynomiale avec des noeuds.

Soit $R \in W_{m,n}$ et soit $Z = \{z_1, \dots, z_l\}$ un ensemble de points de $[a, b]$. On suppose que ces points ne sont pas des noeuds et vérifient les conditions :

$$y_R(z_i) \neq 0 \quad i = 1, \dots, l$$

$$a \leq z_1 < z_2 < \dots < z_l \leq b$$

On considère l'ensemble $S = \{s_1, \dots, s_{l+k}\}$ obtenu en ordonnant la réunion des ensembles Z et $T = \{\xi_1, \dots, \xi_k\}$.

On appelle suite $\tau(R, Z)$ la suite des nombres $\tau_1, \dots, \tau_{l+k}$ définis par :

$$\begin{cases} \tau_j = -\tau_{j-1} & \text{si } s_j \in T \\ \tau_j = \text{signe } y_R(s_j) & \text{si } s_j \in Z \end{cases}$$

où l'on pose :

$$\tau_0 = (-1)^p \text{signe } y_R(z_1) \quad \text{où } z_1 = s_p$$

On démontre tout d'abord le résultat suivant :

Lemme: si la suite $\tau(R, Z)$ présente N changements de signe, il existe au moins un ensemble $Z' \subset Z$ tel que la suite $\tau(R, Z')$ présente N changements de signe et vérifie en outre la condition:

$$\tau_{i+1} = -\tau_i \quad (i = 1, \dots, N)$$

Remarquons que l'ensemble Z comporte au moins $N+1-K$ points. Si Z possède exactement $N+1-K$ points, on prend $Z' = Z$. Sinon il existe un indice i tel que:

$$\tau_{i+1} = \tau_i$$

On voit immédiatement qu'on ne peut pas avoir $s_{i+1} = z_1$ car cela entraînerait:

$$\tau_{i+1} = \text{signe } y_R(z_1)$$

et:

$$\tau_i = (-1)^i \tau_0 = (-1)^{2i+1} \text{signe } y_R(z_1) = -\tau_{i+1}$$

On a donc:

$$s_{i+1} = z_r \quad \text{avec} \quad r \geq 2$$

Supprimons z_r dans l'ensemble Z . La suite $\tau(R, Z - \{s_{i+1}\})$ s'obtiendra simplement en supprimant τ_{i+1} dans la suite $\tau(R, Z)$.

La suite $\tau(R, Z - \{s_{i+1}\})$ présente donc encore N changements de signe.

En répétant cette opération un nombre fini de fois, on aboutira à un ensemble Z' de $N+1-K$ points tel que la suite $\tau(R, Z')$ présente N changements de signe. On aura bien alors:

$$\tau_{i+1} = -\tau_i \quad (i = 1, \dots, N)$$

M étant un sous-ensemble quelconque de $C[a, b]$, nous poserons:

$\nu^*(M) = 1 + \text{nombre maximum de zéros dans } [a, b] \cup \{\xi_1, \dots, \xi_K\}$ des éléments $\Phi \neq 0$ de $\mathcal{D}_R$, en comptant deux fois les zéros intérieurs à $[a, b]$ où Φ ne change pas de signe.

M étant un sous-espace de dimension finie de $C[a, b]$, nous poserons:

$\eta^*(M) = \max \{ \dim(H) \}$, où H est un sous-espace de M tel que:

$$\nu^*(H) \leq K + \dim(H)$$

Tout d'abord, le théorème de La Vallée-Poussin (théorème 3) peut se mettre sous la forme suivante :

Théorème 7 : soit $R \in W_{m,n}$. S'il existe un ensemble $Z = \{z_1, \dots, z_N\}$ de points de $[a, b]$, distincts des noeuds, tels que $y_R(z_i) \neq 0$ et si la suite $\tau(R, Z)$ présente au moins $\nu^*(\mathcal{D}_R)$ changements de signe, alors :

$$\rho \geq \min_{1 \leq i \leq N} |f(z_i) - R(z_i)|$$

D'après le lemme, on peut supposer que les éléments de la suite $\tau(R, Z)$ vérifient :

$$\tau_{i+1} = -\tau_i \quad \forall i$$

La suite $\tau(R, Z)$ comporte alors $N+K$ éléments et donc :

$$N+K-1 \geq \nu^*(\mathcal{D}_R)$$

D'après le théorème 3, il suffit de vérifier qu'il n'existe aucun élément $\Phi \in \mathcal{D}_R$ vérifiant les inégalités :

$$(7) \quad \Phi(x_i) y_R(x_i) > 0 \quad i = 1, \dots, N$$

Supposons qu'un tel élément existe. Formons l'ensemble $S = \{s_1, \dots, s_{l+K}\}$ obtenu en ordonnant la réunion des ensembles Z et $T = \{\xi_1, \dots, \xi_K\}$.

Soient s_i et s_{i+j+1} ($j \geq 0$) deux points de Z consécutifs.

On a :

$$\tau_{i+j+1} = (-1)^{j+1} \tau_i$$

D'où, d'après (7) :

$$(-1)^{j+1} \Phi(s_i) \Phi(s_{i+j+1}) > 0$$

Le nombre de zéros de Φ dans $]s_i, s_{i+j+1}[$, en comptant deux fois les zéros où Φ ne change pas de signe, a donc même parité que $(j+1)$. Comme il y a déjà j zéros dans cet intervalle (les noeuds), il y en a donc au moins $(j+1)$.

Ce qui montre que dans l'ensemble $[a, b] \cup \{\xi_1, \dots, \xi_K\}$, $\Phi(x)$ a au moins $N+K-1$ zéros

Or :

$$N+K-1 \geq \nu^*(\mathcal{D}_R)$$

ce qui, d'après la définition de ν^* , montre que $\Phi(x) \equiv 0$, en contradiction avec la relation (7).

Donc Φ n'existe pas et le théorème 3 nous donne le résultat cherché.

Du théorème 7, on déduit immédiatement une condition suffisante pour que $R \in W_{m,n}$ soit optimal :

Théorème 8 : si $R \in W_{m,n}$ vérifie l'une des conditions suivantes, alors R est optimal pour f :

$$(c1) \quad \|f - R\| = \max_{\xi_i \in [a,b]} |f(\xi_i) - \theta_i|$$

(c2) Il existe un ensemble Z de (e) -points de $f - R$, distincts des noeuds, tels que la suite $\tau(R, Z)$ présente au moins $v^*(\mathcal{D}_R)$ changements de signe.

(c3) Il existe au moins $[v^*(\mathcal{D}_R) + 1 - K]$ (e) -points, distincts des noeuds, tels qu'entre deux (e) -points consécutifs de même genre, il y a un nombre impair de noeuds et entre deux (e) -points consécutifs de genres différents, un nombre pair (ou nul) de noeuds.

Pour (c1), c'est évident. Pour (c2), cela résulte du théorème 7.

Enfin, il est facile de voir que les conditions (c2) et (c3) sont équivalentes.

Réciproquement, si R est optimal, nous allons démontrer le résultat suivant :

Théorème 9 : si $R \in W_{m,n}$ est optimal pour f , l'une des conditions suivantes est vérifiée :

$$(c1) \quad \|f - R\| = \max_{\xi_i \in [a,b]} |f(\xi_i) - \theta_i|$$

(c2) Il existe un ensemble Z de (e) -points, distincts des noeuds, tels que la suite $\tau(R, Z)$ présente au moins $\eta^*(\mathcal{D}_R) + K$ changements de signe.

(c3) Il existe au moins $[1 + \eta^*(\mathcal{D}_R)]$ (e) -points de $f - R$, distincts des noeuds, possédant la propriété (c3) du théorème précédent.

Soit $R \in W_{m,n}$ optimal pour f et supposons :

$$\|f - R\| > \max_{\xi \in [a,b]} |f(\xi_i) - \theta_i|$$

Nous allons montrer que la condition (c3) est vérifiée, cette condition étant comme précédemment équivalente à la condition (c2).

Soit M un sous-espace de $\mathcal{D}_R$ de dimension $N = \eta^*(\mathcal{D}_R)$ vérifiant la condition : $\nu^*(M) \leq K + N$ et soit $\Phi_1(x), \dots, \Phi_N(x)$ une base de M .

M étant contenu dans $\mathcal{D}_R$, d'après le théorème 4, il n'existe aucun élément $\Phi \in M$ tel que :

$$\Phi(x) [f(x) - R(x)] > 0 \quad \forall x \in E_R$$

c'est-à-dire que le système linéaire suivant ne comporte aucune solution $(c_1, \dots, c_N)$:

$$(8) \quad \sum_{i=1}^N c_i \Phi_i(x) Y_R(x) > 0 \quad \forall x \in E_R$$

Donc (théorème 0-2), l'origine de R^N appartient à l'enveloppe convexe de l'ensemble de points :

$$\{[Y_R(x) \Phi_1(x), \dots, Y_R(x) \Phi_N(x)] \in R^N \mid x \in E_R\}$$

Donc (théorème 0-1), il existe $r \leq N+1$ points $x_1, \dots, x_r$ de E_R tels que l'origine de R^N appartienne à l'enveloppe convexe de l'ensemble de points :

$$(9) \quad \{[Y_R(x_i) \Phi_1(x_i), \dots, Y_R(x_i) \Phi_N(x_i)] \mid i = 1, \dots, r\}$$

Remarquons que E_R comporte au moins $(N+1)$ points. Sinon on aurait :

$$E_R = \{x_1, \dots, x_p\} \quad (p \leq N)$$

On pourrait compléter E_R par des points $x_{p+1}, \dots, x_N$ (lorsque $p < N$) n'appartenant pas à E_R et distincts des noeuds.

La relation : $\nu^*(M) \leq K + N$ montre alors que l'on a :

$$\begin{vmatrix} \Phi_1(x_1) & \dots & \Phi_N(x_1) \\ \dots & \dots & \dots \\ \Phi_1(x_N) & \dots & \Phi_N(x_N) \end{vmatrix} \neq 0$$

ce qui prouve qu'il existe des solutions du système (8).

Il y a contradiction et donc E_R comporte bien au moins $(N+1)$ points.

Dans la formule (9), nous pouvons donc prendre :

$$r = N + 1 .$$

De la formule (9), on déduit, en utilisant le théorème 0-5, le résultat suivant :

Le nombre de noeuds contenus entre deux (e)-points consécutifs x_p et x_{p+1} est :

$$\text{impair si} \quad y_R(x_p) y_R(x_{p+1}) > 0$$

$$\text{pair (ou nul) si} \quad y_R(x_p) y_R(x_{p+1}) < 0$$

Finalement on a donc obtenu $N+1 = \eta^*(\mathcal{D}_R) + 1$ points de E_R vérifiant la condition (c3), ce qui termine la démonstration du théorème.

5- Unicité.

Les théorèmes d'unicité du chapitre précédent (théorèmes III-11, 12 et 13) ne donnent aucun résultat intéressant dans le cas présent car il est facile de voir que, dès qu'il y a un noeud ξ_i dans X , $R(x)$ ne possède plus la propriété (P) qui intervient dans ces théorèmes.

Nous allons démontrer des théorèmes d'unicité légèrement différents valables dans le cas où les contraintes sont des conditions de noeuds et nous vérifierons ultérieurement que, dans le cas où tous les noeuds sont extérieurs à X , ces résultats sont identiques à ceux du chapitre précédent.

Nous nous placerons tout d'abord dans le cas général de l'approximation sur un ensemble X compact, pour donner un résultat analogue au théorème III-11.

Désignons par $\tilde{C}^*[X]$ l'ensemble des fonctions $f \in C[X]$ telles que :

a) $f(x)$ a au moins un élément optimal R dans $W_{m,n}$.

b) $\|f - R\| > \max_{\xi_i \in X} |f(\xi_i) - \theta_i|$.

On a le résultat suivant :

Théorème 10 : Toute fonction $f \in \tilde{C}^*[X]$ a un élément optimal unique dans $W_{m,n}$ si et seulement si toute fraction $R \in W_{m,n}$ possède la propriété (P1) ci-dessous.

Propriété (P1) : il n'existe pas de points $x_1, \dots, x_r$ de I , distincts des noeuds, et de nombres λ_i, γ_i tels que :

$$(10) \quad \lambda_i \neq 0 \quad 1 \leq r \leq \dim(\mathcal{D}_R)$$

$$(11) \quad \sum_{i=1}^r \lambda_i \begin{bmatrix} g_0(x_i) \\ \vdots \\ g_m(x_i) \\ -R(x_i)h_0(x_i) \\ \vdots \\ -R(x_i)h_n(x_i) \end{bmatrix} + \sum_{i=1}^K \gamma_i \begin{bmatrix} g_0(\xi_i) \\ \vdots \\ g_m(\xi_i) \\ -\theta_i h_0(\xi_i) \\ \vdots \\ -\theta_i h_n(\xi_i) \end{bmatrix} = 0$$

Remarque : comme on l'a déjà vu, la condition (11) est équivalente à la condition suivante :

$$\sum_{i=1}^r \lambda_i \Phi(x_i) = 0 \quad \forall \Phi \in \mathcal{D}_R$$

La démonstration du théorème 10 va être à peu près l'analogue de celle du théorème II-6 :

1°) Soit $R \in W_{m,n}$ optimal pour $f \in \tilde{C}^*[X]$. Supposons que R possède la propriété (P1) et montrons qu'alors R est le seul élément optimal pour f . Supposons au contraire que f possède dans $W_{m,n}$ un autre élément optimal $R^*(x)$ distinct de $R(x)$ et posons :

$$R(x) = \frac{\sum_0^m a_i g_i(x)}{\sum_0^n b_i h_i(x)}, \quad R^*(x) = \frac{\sum_0^m a_i^* g_i(x)}{\sum_0^n b_i^* h_i(x)}$$

Posons $N = \dim(\mathcal{D}_R)$ et soit $\Phi_1(x), \dots, \Phi_N(x)$ une base de $\mathcal{D}_R$.

$R(x)$ étant optimal, il existe, d'après le théorème 5, des (e)-points $x_1, \dots, x_r$ de $f-R$ et des nombres $\lambda_1, \dots, \lambda_r$ tels que :

$$(13) \quad \begin{cases} \lambda_i [f(x_i) - R(x_i)] > 0 \\ 1 \leq r \leq N+1 \\ \sum_{i=1}^r \lambda_i \Phi_j(x_i) = 0 \end{cases} \quad j=1, \dots, N$$

En retranchant (14) de (13), nous obtenons alors :

$$\lambda_1 \begin{bmatrix} \Phi_1(x_1) \\ \vdots \\ \vdots \\ \vdots \\ \Phi_N(x_1) \end{bmatrix} + \sum_{\substack{i=2 \\ i \neq i_0}}^{N+1} (\lambda_i - \lambda_{i'}) \begin{bmatrix} \Phi_1(x_i) \\ \vdots \\ \vdots \\ \vdots \\ \Phi_N(x_i) \end{bmatrix} = 0$$

Ceci est en contradiction avec le fait que $R(x)$ possède la propriété (P1). Donc $R^i(x)$ ne peut pas exister et $R(x)$ est bien le seul élément optimal pour f dans $W_{m,n}$.

2°) Réciproquement nous allons montrer que s'il existe une fonction $R(x)$ ne possédant pas la propriété (P1), on peut construire une fonction $f \in \tilde{C}^*[X]$ qui possède deux éléments optimaux distincts dans $W_{m,n}$.

Supposons donc qu'il existe une fraction :

$$R(x) = \frac{\sum_0^m a_j^* g_j(x)}{\sum_0^n b_j^* h_j(x)}$$

telle que :

$$(15) \left\{ \begin{array}{ll} \sum_{j=0}^m a_j^* g_j(\xi_i) - \theta_i \sum_{j=0}^n b_j^* h_j(\xi_i) = 0 & i = 1, \dots, K \\ \sum_{j=0}^n b_j^* h_j(x) > 0 & \text{sur } X \\ \lambda_i \neq 0 & 1 \leq r \leq N \\ \sum_{i=1}^r \lambda_i \Phi_j(x_i) = 0 & j = 1, \dots, N \end{array} \right.$$

(a) Montrons qu'il existe dans $\mathcal{D}_R$ un élément $\Phi(x)$ non identiquement nul tel que :

$$\Phi(x_i) = 0 \quad i = 1, \dots, r$$

Il nous faut donc résoudre le système linéaire :

$$(16) \quad \sum_{j=1}^N \tau_j \Phi_j(x_i) = 0 \quad i = 1, \dots, r$$

Mais d'après (15), on a :

$$\sum_{i=1}^r \lambda_i \begin{bmatrix} \Phi_1(x_i) \\ \vdots \\ \Phi_N(x_i) \end{bmatrix} = 0$$

ce qui prouve que la matrice du système (16) est au plus de rang :

$$r-1 \leq N-1$$

comme (16) est un système homogène à N inconnues, ceci prouve qu'il existe une solution non identiquement nulle de ce système.

(b) Désignons donc par $\Phi(x)$ un élément de $\mathcal{D}_R$ non identiquement nul :

$$\Phi(x) = \sum_0^m a_j g_j(x) - R(x) \sum_0^n b_j h_j(x)$$

$$\text{tel que :} \quad \Phi(x_i) = 0 \quad i = 1, \dots, r$$

Considérons alors la fraction suivante :

$$R_1(x) = \frac{\sum_0^m a_j^* g_j(x)}{\sum_0^n b_j^* h_j(x)}$$

avec :

$$\begin{aligned} a_j^* &= \theta a_j + (1-\theta) a_j^* & j &= 0, 1, \dots, m \\ b_j^* &= \theta b_j + (1-\theta) b_j^* & j &= 0, 1, \dots, n \\ 0 &< \theta < 1 \end{aligned}$$

Pour θ assez petit, le dénominateur est strictement positif sur X et alors $R_1 \in W_{m,n}$.

De plus, on a :

$$R(x) - R_1(x) = \frac{-\theta \Phi(x)}{\sum_0^n b_j^* h_j(x)}$$

Ceci montre que $R(x)$ et $R_1(x)$ sont deux éléments distincts de $W_{m,n}$ qui vérifient en outre :

$$R(x_i) = R_1(x_i) \quad i = 1, \dots, r$$

c'est-à-dire que la fraction $R_1(x)$ vérifie les relations (15).

Enfin, on peut choisir θ assez petit pour que

$$\|R - R_1\| \leq 1$$

(c) On pose :

$$f(x) = R(x) + [1 - |R(x) - R_1(x)|] g(x)$$

avec :

$$g \in C[X] \quad , \quad \|g\| = 1$$

$$g(x_i) = \text{signe } (\lambda_i) \quad i = 1, \dots, r$$

$$|g(\xi_i)| < 1 \quad \text{si} \quad \xi_i \notin X$$

On montre alors (voir démonstration du théorème II-6) que l'on a :

$$\|f - R\| = \|f - R_1\| = 1$$

De plus, les points $x_1, \dots, x_r$ sont (θ) -points de $f - R$ et de $f - R_1$ avec :

$$\lambda_i [f(x_i) - R(x_i)] = \lambda_i [f(x_i) - R_1(x_i)] > 0$$

Comme R et R_1 vérifient les relations (15), ceci prouve que R et R_1 sont deux éléments optimaux distincts pour f dans $W_{m,n}$.

En outre, la condition: $|g(\xi_i)| < 1$ montre que :

$$\|f - R\| > \max_{\xi_i \in X} |f(\xi_i) - \theta_i|$$

c'est-à-dire que $f \in \tilde{C}^*[X]$.

Nous avons donc bien construit une fonction de $\tilde{C}^*[X]$ possédant dans $W_{m,n}$ deux éléments optimaux distincts, ce qui termine la démonstration du théorème 10.

Dans le cas particulier où $X = [a, b]$ intervalle fini de la droite réelle, nous allons démontrer à l'aide des résultats du paragraphe précédent un résultat analogue au théorème III-12.

Nous désignerons par $\tilde{X}$ l'intervalle $[a, b]$ privé des noeuds qu'il contient et nous dirons que $\mathcal{D}_R$ vérifie la condition de Haar sur $\tilde{X}$ si tout élément $\phi(x)$ de $\mathcal{D}_R$ non identiquement nul a au plus $\dim(\mathcal{D}_R) - 1$ zéros dans $\tilde{X}$. Ceci est équivalent, en désignant par $\phi_1(x), \dots, \phi_N(x)$

une base de $\mathcal{D}_R$ à la condition :

$$\begin{vmatrix} \Phi_1(x_1) & \dots & \Phi_1(x_N) \\ \dots & \dots & \dots \\ \Phi_N(x_1) & \dots & \Phi_N(x_N) \end{vmatrix} \neq 0$$

quels que soient $x_1, \dots, x_N$ distincts dans $\tilde{X}$.

Remarquons que si $\nu^*(\mathcal{D}_R) \leq K + \eta^*(\mathcal{D}_R)$, alors tout élément $\Phi(x) \neq 0$ de $\mathcal{D}_R$ a au plus :

$$\eta^*(\mathcal{D}_R) - 1 \leq \dim(\mathcal{D}_R) - 1$$

zéros dans $\tilde{X}$ et donc $\mathcal{D}_R$ vérifie la condition de Haar sur $\tilde{X}$ (la réciproque étant fautive à cause de la présence possible de noeuds intérieurs à $\tilde{X}$ comme zéros doubles).

On a alors le résultat suivant :

Théorème 11 : Toute fonction $f \in \tilde{C}^*[a, b]$ a un élément optimal unique dans $W_{m, n}$ si et seulement si, pour tout $R \in W_{m, n}$, $\mathcal{D}_R$ vérifie la condition de Haar sur $\tilde{X}$.

La démonstration de ce théorème est analogue à celle du théorème III-12 avec quelques légères modifications (dans la réciproque, la fonction $f \in \tilde{C}^*[a, b]$ possédant deux éléments optimaux distincts doit être définie comme dans le (c) de la réciproque du théorème 10).

Le théorème peut aussi se déduire du théorème 10 à l'aide du lemme suivant (dont la démonstration est analogue à celle du lemme utilisé dans la remarque 1 du théorème III-12) :

Lemme : lorsque $X = [a, b]$, $R \in W_{m, n}$ possède la propriété (P1) si et seulement si $\mathcal{D}_R$ vérifie la condition de Haar sur $\tilde{X}$.

Enfin la première partie de la démonstration du théorème 10 nous fournit une condition suffisante d'unicité qui s'énonce :

Théorème 12 : soit $R \in W_{m, n}$ optimal pour $f \in C[X]$. Si R possède la propriété (P1) (ou, lorsque $X = [a, b]$, si $\mathcal{D}_R$ vérifie la condition de

Haar sur $\tilde{X}$) et si :

$$\|f - R\| > \max_{\xi_i \in X} |f(\xi_i) - \theta_i| ,$$

alors R est le seul élément optimal pour f dans $W_{m,n}$.

6 - Unicité forte.

Dans ce paragraphe, nous supposerons encore $X = [a, b]$.

Soit R^* optimal pour f dans $W_{m,n}$. Le théorème 12 montre que les hypothèses :

$$\nu^*(\mathcal{D}_{R^*}) \leq \eta^*(\mathcal{D}_{R^*}) + K$$

$$\|f - R^*\| > \max_{\xi_i \in [a, b]} |f(\xi_i) - \theta_i|$$

entraînent :

$$\|f - R\| - \|f - R^*\| > 0 \quad \forall R \in W_{m,n}, \quad R \neq R^*$$

Nous allons démontrer que, si l'on a en outre :

$$\dim(\mathcal{D}_{R^*}) = m + n + 1 - K$$

un résultat analogue au théorème III-14 est valable, c'est-à-dire qu'il existe une constante γ strictement positive telle que :

$$\|f - R\| - \|f - R^*\| > \gamma \|R - R^*\|$$

Ce résultat s'énonce sous la forme suivante :

Théorème 13 : soit R^* optimal pour f dans $W_{m,n}$.

Si $\eta^*(\mathcal{D}_{R^*}) + K = \nu^*(\mathcal{D}_{R^*}) = m + n + 1$ et si :

$$\|f - R\| > \max_{\xi_i \in [a, b]} |f(\xi_i) - \theta_i| ,$$

alors il existe une constante $\gamma > 0$ telle que :

$$\|f - R\| \geq \|f - R^*\| + \gamma \|R - R^*\|$$

La démonstration suit dans ses grandes lignes celle de théorème III-14. La seule modification porte sur la vérification de la propriété suivante :

si $\Phi(x) \in \mathcal{D}_{R^*}$ est tel que :

$$\Phi(x) [f(x) - R^*(x)] \geq 0 \quad \forall x \in E_{R^*},$$

alors on a :

$$\Phi(x) \equiv 0$$

Nous allons le démontrer.

$R^*(x)$ étant optimal, il existe (d'après le théorème 9) un ensemble Z de (e) -points, distincts des noeuds tels que la suite $\tau(R^*, Z)$ présente un nombre N de changements de signe au moins égal à :

$$\eta^*(\mathcal{D}_{R^*}) + K = m + n + 1$$

On peut supposer (d'après le lemme du paragraphe 4) que la suite $\tau(R^*, Z)$ vérifie la condition :

$$\tau_{i+1} = -\tau_i \quad i = 1, \dots, N$$

Posons $Z = \{z_1, \dots, z_{N-K+1}\}$ et formons l'ensemble $S = \{s_1, \dots, s_{N+1}\}$ obtenu en ordonnant la réunion de Z et de l'ensemble $\{\xi_1, \dots, \xi_K\}$

On a :

$$\tau_i = (-1)^i \varepsilon \quad (\text{où } \varepsilon = \pm 1)$$

On a d'autre part :

$$\begin{aligned} \Phi(s_j) &= 0 & \text{si } s_j \in \{\xi_1, \dots, \xi_K\} \\ \Phi(s_j)[f(s_j) - R^*(s_j)] &\geq 0 & \text{si } s_j \in Z \end{aligned}$$

c'est-à-dire :

$$(-1)^j \varepsilon \Phi(s_j) \geq 0 \quad j = 1, \dots, N+1$$

Soient alors s_j et s_{j+r+1} ($r \geq 0$) deux points consécutifs où Φ ne s'annule pas. On a alors :

$$(-1)^{r+1} \Phi(s_j) \Phi(s_{j+r+1}) > 0$$

Donc en comptant deux fois les zéros intérieurs à $[a, b]$ où $\Phi(x)$ ne change pas de signe, $\Phi(x)$ a au moins $(r+1)$ zéros dans l'intervalle $]s_j, s_{j+r+1}[$.

Au total, $\Phi(x)$ a donc au moins N zéros (en comptant les multiplisités) dans l'ensemble $[a, b] \cup \{\xi_1, \dots, \xi_K\}$.

Comme d'autre part, on a :

$$\nu^*(\mathcal{D}_{R^*}) = m + n + 1 \leq N$$

ceci prouve que :

$$\Phi(x) \equiv 0$$

Remarque : l'hypothèse $\eta^*(\mathcal{D}_R) + K = \nu^*(\mathcal{D}_R) = m + n + 1$ peut être remplacée par :

$$\begin{cases} \nu^*(\mathcal{D}_R) \leq K + \eta^*(\mathcal{D}_R) \\ \dim(\mathcal{D}_R) = m + n + 1 - K \end{cases}$$

En effet, la première inégalité montre que $\mathcal{D}_R$ vérifie la condition de Haar sur $\tilde{X}$ et donc :

$$\eta^*(\mathcal{D}_R) = \dim(\mathcal{D}_R) = m + n + 1 - K = M$$

D'autre part, si $x_1, \dots, x_M$ sont des points distincts dans $\tilde{X}$ et si $\Phi_1(x), \dots, \Phi_M(x)$ est une base de $\mathcal{D}_R$, le système linéaire suivant a une solution :

$$\begin{aligned} \sum_{j=1}^M \tau_j \Phi_j(x_i) &= 0 & i = 1, \dots, M-1 \\ \sum_{j=1}^M \tau_j \Phi_j(x_M) &= 1 \end{aligned}$$

Ceci prouve qu'il existe dans $\mathcal{D}_R$ un élément $\Phi \neq 0$ qui admet $M-1$ zéros en dehors des noeuds. D'où :

$$\nu^*(\mathcal{D}_R) \geq M + K = m + n + 1$$

et en tenant compte de la première inégalité :

$$\nu^*(\mathcal{D}_R) = m + n + 1$$

7- Continuité.

Dans ce paragraphe également, nous supposons $X = [a, b]$.

Etant donné $f \in C[a, b]$, nous noterons $\mathcal{T}_f$ l'ensemble des fractions de $W_{m,n}$ qui réalisent la meilleure approximation de f .

Nous allons démontrer un résultat analogue au théorème III-15 et nous vérifierons ultérieurement qu'il y a identité entre ces deux théorèmes lorsque tous les noeuds sont extérieurs à $[a, b]$.

Théorème 14: si $R_0 \in \mathcal{T}f_0$, $\eta^*(\mathcal{D}_{R_0}) + K = \nu^*(\mathcal{D}_{R_0}) = m + n + 1$ et si :

$$\|f_0 - R_0\| > \max_{\xi_i \in [a,b]} |f_0(\xi_i) - \theta_i| ,$$

alors $\mathcal{T}f$ est non vide pour tous les f d'un voisinage de f_0 et il existe $\beta > 0$ tel que :

$$\|R - R_0\| < \beta \|f - f_0\| \quad \forall R \in \mathcal{T}f$$

La démonstration de ce résultat est identique à celle du théorème III-15.

8- Cas des noeuds extérieurs à l'intervalle $[a,b]$.

Dans le cas particulier où $X = [a,b]$ et où tous les noeuds sont extérieurs à l'intervalle $[a,b]$, nous allons comparer ce qu'on obtient en appliquant les théorèmes du présent chapitre et ceux du chapitre précédent.

En ce qui concerne la caractérisation des éléments optimaux, le théorème IV-4 devient identique au théorème III-4.

En ce qui concerne les propriétés d'alternance, il nous faut tout d'abord calculer les nombres $\nu(M)$, $\eta(M)$, $\nu^*(M)$, $\eta^*(M)$.

On a :

$$\nu(M) = 1 + \text{nombre maximum de changements de signe dans } [a,b] \text{ des éléments } \Phi \neq 0 \text{ de } M$$

et :

$$\nu^*(M) = K + 1 + \text{nombre maximum de zéros dans } [a,b] \text{ des éléments } \Phi \neq 0 \text{ de } M , \text{ en comptant deux fois les zéros intérieurs à } [a,b] \text{ où } \Phi(x) \text{ ne change pas de signe.}$$

On en déduit :

$$(17) \quad \nu^*(M) \geq K + \nu(M)$$

D'autre part :

$$\eta(M) = \max \{ \dim(H) , \text{ où } H \text{ est un sous-espace de Haar de } M \}$$

et :

$$\eta^*(M) = \max \{ \dim(H) , \text{ où } H \text{ est un sous-espace de } M \text{ tel que : } \nu^*(H) \leq K + \dim(H) \}$$

Or, nous avons déjà vu que si :

$$\nu^*(H) \leq K + \dim(H)$$

alors H est un sous-espace de Haar (dans le cas où tous les noeuds sont extérieurs à l'intervalle).

Réciproquement, si H est un sous-espace de Haar, tout élément $\Phi \neq 0$ de H a dans $[a, b]$ au plus $\dim(H) - 1$ zéros, en comptant les multiplicités (théorème 0-3), et donc alors :

$$\nu^*(H) \leq K + \dim(H)$$

Ce qui montre que :

$$(18) \quad \eta^*(M) = \eta(M)$$

En outre, on a le résultat suivant :

$$(19) \quad H \text{ est un sous espace de Haar} \iff \nu^*(H) = \eta^*(H) + K$$

En effet, si on a :

$$\nu^*(H) = \eta^*(H) + K$$

ceci entraîne d'après ce que nous venons de voir que H est un sous-espace de Haar.

Réciproquement, si H est un sous-espace de Haar, on a :

$$\dim(H) = \nu(H) = \eta(H)$$

et d'autre part :

$$\nu^*(H) \leq K + \dim(H)$$

D'où d'après (17) et (18) :

$$\nu^*(H) = K + \eta^*(H)$$

Nous pouvons maintenant comparer les résultats des chapitres III et IV.

Théorème de La Vallée-Poussin : le théorème III-8 semble meilleur que le théorème IV-7 (ces résultats étant équivalents si $\mathcal{D}_R$ est un sous-espace de Haar).

En effet, le théorème III-8 demande $1 + \nu(\mathcal{D}_R)$ points et le théorème IV-7 en demande :

$$1 + \nu^*(\mathcal{D}_R) - K \geq \nu(\mathcal{D}_R) + 1$$

Et dans le cas où $\mathcal{D}_R$ est un sous-espace de Haar, on a d'après (19) :

$$1 + \nu^*(\mathcal{D}_R) - K = 1 + \eta^*(\mathcal{D}_R) = 1 + \eta(\mathcal{D}_R) = 1 + \nu(\mathcal{D}_R)$$

Propriétés d'alternance : le théorème III-9 semble meilleur que le théorème IV-8 (ces résultats sont équivalents si $\mathcal{D}_R$ est un sous espace de Haar). D'autre part, les théorèmes III-10 et IV-9 deviennent identiques.

Cela résulte de ce qui précède et de la relation (18).

Unicité : les théorèmes III-11, 12 et 13 deviennent identiques aux théorèmes IV-10, 11 et 12.

En effet, les classes de fonctions $\tilde{C}^*[a,b]$ et $\tilde{C}[a,b]$ coïncident et les propriétés (P) et (P1) sont identiques.

Unicité forte et continuité : les théorèmes III-14, 15 deviennent identiques aux théorèmes IV-13, 14.

En effet, l'hypothèse $\eta(\mathcal{D}_R) = m+n+1-K$ est équivalente à : $\mathcal{D}_R$ sous-espace de Haar de dimension $m+n+1-K$ et il en est de même pour l'hypothèse :

$$\eta^*(\mathcal{D}_R) + K = \nu^*(\mathcal{D}_R) = m+n+1$$

(voir remarque du théorème 13).

9- Retour sur le cas des contraintes inégalités.

Revenons à un cas particulier du problème considéré dans le chapitre II. Considérons l'approximation d'une fonction $f \in C[a,b]$ par la classe $W_{m,n}$ des fractions rationnelles :

$$R(x) = \frac{\sum_0^m a_j g_j(x)}{\sum_0^n b_j h_j(x)}$$

dont les coefficients a_j, b_j vérifient les contraintes suivantes :

$$u(x) \leq R(x) \leq V(x) \quad \text{sur} \quad [a, b]$$

$$\sum_0^n b_j h_j(x) > 0 \quad \text{sur} \quad [a, b]$$

où $u(x)$ et $V(x)$ sont deux fonctions données dans $C[a, b]$.

Remarquons que la première condition s'écrit encore :

$$\begin{aligned} \sum_0^m a_j g_j(t) - u(t) \sum_0^n b_j h_j(t) &\geq 0 & \forall t \in [a, b] \\ \sum_0^m -a_j g_j(t) + V(t) \sum_0^n b_j h_j(t) &\geq 0 & \forall t \in [a, b] \end{aligned}$$

Soit $R(x)$ optimal pour f dans $W_{m,n}$.

Les théorèmes II-5 (remarque 3) et IV-5 montrent qu'il existe des points $t_1, \dots, t_s, t_{s+1}, \dots, t_{s+s'}$ ($s+s' \geq 0$) de $[a, b]$ tels que $R(x)$ soit optimal pour f dans la classe $V(t_1, \dots, t_s, t_{s+1}, \dots, t_{s+s'})$ des fractions rationnelles :

$$R(x) = \frac{\sum_0^m a_j g_j(x)}{\sum_0^n b_j h_j(x)}$$

dont les coefficients a_j, b_j vérifient les contraintes suivantes :

$$\sum_0^n b_j h_j(x) > 0 \quad \text{sur} \quad [a, b]$$

$$(20) \quad R(t_i) = u(t_i) \quad i = 1, \dots, s$$

$$(21) \quad R(t_i) = V(t_i) \quad i = s+1, \dots, s+s'.$$

Désignons par $F(R; t_1, \dots, t_s, t_{s+1}, \dots, t_{s+s'})$ la classe des fonctions :

$$\Phi(x) = \sum_0^m a_j g_j(x) - R(x) \sum_0^n b_j h_j(x)$$

dont les coefficients a_j, b_j vérifient les relations :

$$(22) \quad \sum_{j=0}^m a_j g_j(t_i) - u(t_i) \sum_{j=0}^n b_j h_j(t_i) = 0 \quad i = 1, \dots, s$$

$$(23) \quad \sum_{j=0}^m a_j g_j(t_i) - V(t_i) \sum_{j=0}^n b_j h_j(t_i) = 0 \quad i = s+1, \dots, s+s'$$

Aux résultats (A) et (B) du chapitre II (paragraphe 5), le théorème 9 permet d'ajouter le résultat suivant :

(C) Soit $R(x)$ optimal pour f dans $W_{m,n}$.

(1) Il existe $s+s'$ points $t_1, \dots, t_{s+s'}$ de $[a, b]$ tels que :

$$R(t_i) = u(t_i) \quad i = 1, \dots, s$$

$$R(t_i) = V(t_i) \quad i = s+1, \dots, s+s'$$

(2) Si aucun des points t_i n'est (e)-point de $f-R$, alors il existe au moins $1+\eta^* [F(R; t_1, \dots, t_s, t_{s+1}, \dots, t_{s+s'})]$ (e)-points de $f-R$, distincts des points t_i tels qu'entre deux (e)-points consécutifs de même genre il y a un nombre impair de points t_i et entre deux (e)-points consécutifs de genre différents un nombre pair (ou nul) de points t_i .

Enfin, le théorème 12 permet de donner une autre forme du théorème II-7 :

(D) Soit $R \in W_{m,n}$ optimal pour f . Si $F(R; t_1, \dots, t_{s+s'})$ vérifie la condition de Haar sur $[a, b] - \{t_1, \dots, t_{s+s'}\}$ et si $\|f-R\| > m_{t_i} |f(t_i) - R(t_i)|$, alors R est le seul élément optimal pour f dans $W_{m,n}$.

En effet, d'après le théorème 12, R est le seul élément optimal pour f dans la classe $V(t_1, \dots, t_{s+s'})$ définie au début de ce paragraphe. On vérifie ensuite (comme au paragraphe 8 du chapitre précédent) que f ne peut pas admettre un autre élément optimal $R' \in W_{m,n}$ car un tel élément appartiendrait aussi à $V(t_1, \dots, t_{s+s'})$.

Chapitre V

ETUDE DE QUELQUES CAS PARTICULIERS

Dans ce chapitre, nous allons étudier quelques cas particuliers des problèmes généraux précédemment envisagés, en insistant surtout sur l'existence et les propriétés d'alternance.

Dans le paragraphe 1, on étudie l'approximation par des fractions rationnelles "ordinaires"

$$R(x) = \frac{a_0 + a_1 x + \dots + a_m x^m}{b_0 + b_1 x + \dots + b_n x^n}$$

prenant des valeurs données en des points donnés.

Dans le paragraphe 2, on étudie le même problème pour des fractions rationnelles "trigonométriques"

$$R(\theta) = \frac{\sum_0^m (a_j \cos j\theta + b_j \sin j\theta)}{\sum_0^n (c_j \cos j\theta + b_j \sin j\theta)}$$

Dans le paragraphe 3, on étudie l'approximation par des fractions rationnelles ordinaires avec des contraintes inégalités du type :

$$V(x) \leq R(x) \leq W(x)$$

et dans le paragraphe 4, on étudie le même problème pour des fractions rationnelles trigonométriques.

1 - Fractions rationnelles ordinaires avec conditions de noeuds.

Problème: Soit $I = [a, b]$ un intervalle compact de la droite réelle. Désignons par $\xi_1, \dots, \xi_K$ des noeuds contenus dans un compact I' contenant I . On cherche à approcher une fonction $f \in C[a, b]$ par la classe $W_{m, n}$

des fonctions

$$R(x) = u(x) \frac{p(x)}{q(x)}$$

où $u(x)$ désigne une fonction réelle donnée, définie sur I' , continue strictement positive sur I et où $p(x)$, $q(x)$ sont deux polynômes algébriques à coefficients réels vérifiant les conditions suivantes :

$$d^0(p) \leq m, \quad d^0(q) \leq n$$

$$q(x) > 0 \quad \text{sur} \quad [a, b]$$

$$q(\xi_i) \neq 0 \quad i = 1, \dots, K$$

$$p(x) \text{ et } q(x) \text{ sont premiers entre eux}$$

$$R(\xi_i) = \theta_i \quad i = 1, \dots, K$$

où $\theta_1, \dots, \theta_K$ sont des nombres réels donnés.

Existence d'un élément optimal : la classe $\overline{W}_{m,n}$ précédemment définie est l'ensemble des fonctions $R \in C[a, b]$ qui vérifient une identité de la forme :

$$(1) \quad q_1(x) R(x) \equiv u(x) p_1(x)$$

où $p_1(x)$ et $q_1(x)$ sont deux polynômes algébriques à coefficients réels tels que :

$$d^0(p_1) \leq m, \quad d^0(q_1) \leq n$$

$$q_1(x) \neq 0$$

$$u(\xi_i) p_1(\xi_i) - \theta_i q_1(\xi_i) = 0 \quad i = 1, \dots, K$$

Pour $R \in \overline{W}_{m,n}$ vérifiant (1), nous écrirons :

$$R = (p_1 | q_1)$$

Le théorème IV-1 montre que l'on a tout d'abord le résultat suivant :

Théorème 1 : toute fonction $f \in C[a, b]$ possède dans $\overline{W}_{m,n}$ au moins un élément optimal

Remarque: Soit $R = (p_1 | q_1)$ un élément de $\overline{W}_{m,n}$ optimal pour f et supposons que l'on ait:

$$q_1(\xi_i) \neq 0 \quad i = 1, \dots, K$$

Alors $R(x)$ peut être identifié avec un élément de $W_{m,n}$. En effet, la relation (1) montre que tout zéro d'ordre $r(r \geq 1)$ de $q_1(x)$ est zéro de $p_1(x)$ d'ordre au moins égal à r . En divisant $p_1(x)$ et $q_1(x)$ par leur plus grand commun diviseur, on obtient deux polynômes $p_2(x)$ et $q_2(x)$ premiers entre eux et tels que:

$$\begin{aligned} q_2(x) &\neq 0 && \text{sur } [a, b] \\ q_2(\xi_i) &\neq 0 && i = 1, \dots, K \\ u(\xi_i) p_2(\xi_i) - \theta_i q_2(\xi_i) &= 0 && i = 1, \dots, K \end{aligned}$$

Ce qui montre que $R(x)$ est un élément de $W_{m,n}$.

Mais si $q_1(x)$ s'annule en un des noeuds, l'élément $R = (p_1 | q_1)$ peut ne pas appartenir à $W_{m,n}$ comme le montre l'exemple suivant (voir LOEB [1968]):

Prenons $I = [0, 1]$ et soit $f(x)$ la fonction définie par:

$$f(x) = \begin{cases} 1 + 6x & \text{si } 0 \leq x \leq \frac{1}{3} \\ 5 - 6x & \text{si } \frac{1}{3} \leq x \leq \frac{2}{3} \\ -3 + 6x & \text{si } \frac{2}{3} \leq x \leq 1 \end{cases}$$

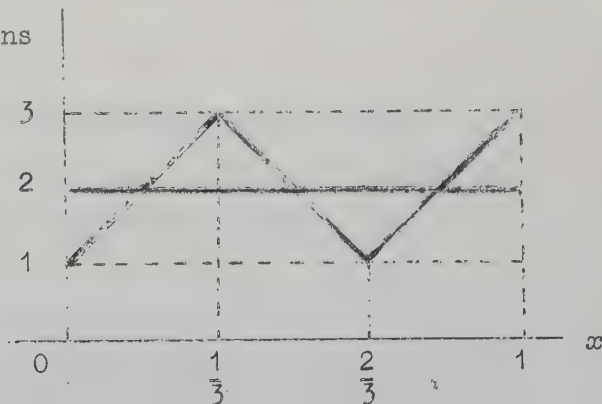
Prenons pour W la classe des fonctions

$$r(x) = \frac{ax + b}{cx + d}$$

telles que:

$$r(0) = 1$$

$$cx + d > 0 \quad \text{sur } [0, 1]$$



La classe $\overline{W}$ est donc l'ensemble des fonctions $R \in C[I]$ qui vérifient une identité de la forme:

$$(\alpha x + \beta) = (\gamma x + \beta) R(x)$$

On vérifie aisément que :

$$\inf_{r \in W} \|f - r\| = \inf_{r \in \bar{W}} \|f - r\| = 1$$

Et la seule fraction qui vérifie $\|f - R\| = 1$ est :

$$R(x) \equiv 2$$

Cette fraction appartient à $\bar{W}_{m,n}$ (en prenant $\alpha = 2$, $\gamma = 1$, $\beta = 0$) mais n'appartient pas à $W_{m,n}$.

Propriétés d'alternance : nous allons calculer les nombres ν^* et η^* définis au chapitre précédent (paragraphe 4).

Soit $R \in W_{m,n}$:

$$R(x) = u(x) \frac{p(x)}{q(x)}$$

Nous poserons : $\bar{d}(R) = m + n + 1 - \min [m - d^0(p), n - d^0(q)]$

Tout élément $\Phi \in \mathcal{D}_R$ s'écrit :

$$\Phi(x) = u(x)(c_0 + c_1 x + \dots + c_m x^m) - u(x) \frac{p(x)}{q(x)}(d_0 + d_1 x + \dots + d_n x^n)$$

avec :

$$\Phi(\xi_i) = 0 \quad i = 1, \dots, K$$

Soit encore :

$$(2) \quad \Phi(x) = \frac{(x - \xi_1) \dots (x - \xi_K) u(x)}{q(x)} \times \Phi_1(x)$$

où $\Phi_1(x)$ est un polynôme arbitraire tel que :

$$d^0(\Phi_1) \leq \bar{d}(R) - 1 - K$$

On déduit immédiatement de la formule (2) :

$$\nu^*(\mathcal{D}_R) = \bar{d}(R)$$

$$\eta^*(\mathcal{D}_R) = \bar{d}(R) - K$$

Théorème de La Vallée-Poussin : compte tenu des relations ci-dessus, le théorème IV-7 devient :

Théorème 2 : soit $R \in W_{m,n}$. S'il existe un ensemble $Z = \{z_1, \dots, z_N\}$ de points de $[a, b]$, distincts des noeuds, tels que $y_R(z_i) \neq 0$ et si la

suite $\tau(R, Z)$ présente au moins $d(R)$ changements de signe, alors on a :

$$\rho \geq \min_{1 \leq i \leq N} |f(z_i) - R(z_i)|$$

Caractérisation des éléments optimaux : compte tenu de la relation :

$$\nu^*(\mathcal{D}_R) = \eta^*(\mathcal{D}_R) + K$$

les théorèmes IV-8 et IV-9 fournissent une condition nécessaire et suffisante pour que $R \in W_{m,n}$ soit optimal :

Théorème 3 : pour que $R \in W_{m,n}$ soit optimal, il faut et il suffit que l'une des conditions suivantes soit vérifiée :

$$(a) \quad \|f - R\| = \max_{\xi_i \in [a,b]} |f(\xi_i) - \theta_i|$$

(b) Il existe un ensemble Z de (e) -points de $f - R$, distincts des noeuds, tels que la suite $\tau(R, Z)$ présente au moins $d(R)$ changements de signe.

Rappelons que la condition (b) est équivalente à la condition suivante : il existe au moins $[d(R) + 1 - K]$ (e) -points de $f - R$, distincts des noeuds, tels qu'entre deux (e) -points consécutifs de même genre il y a un nombre impair de noeuds et entre deux (e) -points consécutifs de genres différents, un nombre pair (ou nul) de noeuds.

Unicité : nous avons vu que pour tout $R \in W_{m,n}$, on a :

$$\nu^*(\mathcal{D}_R) = \eta^*(\mathcal{D}_R) + K$$

ce qui montre que $\mathcal{D}_R$ vérifie la condition de Haar sur $[a,b]$ privé des noeuds. Le théorème IV-12 nous donne donc le résultat suivant :

Théorème 4 : soit $R \in W_{m,n}$ optimal pour f . Si on a :

$$\|f - R\| > \max_{\xi_i \in [a,b]} |f(\xi_i) - \theta_i|$$

alors R est le seul élément optimal pour f dans $W_{m,n}$.

Unicité forte et continuité: Les théorèmes IV-13 et IV-14 subsistent en remplaçant l'hypothèse :

$$\nu^*(\mathcal{D}_{R^*}) = \eta^*(\mathcal{D}_{R^*}) + K = m + n + 1$$

par la seule condition :

$$\bar{d}(R) = m + n + 1$$

Remarque: désignons par $C_0[a, b]$ la classe $\tilde{C}^*[a, b]$ définie au chapitre précédent. C'est la classe des fonctions $f \in C[a, b]$ telles que :

- a) f a au moins un élément optimal $R^* \in W_{m, n}$
- b) cet élément vérifie la condition :

$$\|f - R^*\| > \max_{\xi_i \in [a, b]} |f(\xi_i) - \theta_i|$$

Des théorèmes précédents, on déduit les résultats suivants :

- (A) Toute fonction $f \in C_0[a, b]$ a dans $W_{m, n}$ un élément optimal unique.
- (B) Cet élément est caractérisé par l'existence d'un ensemble Z de (e) -points tel que la suite $\tau(R, Z)$ présente au moins $\bar{d}(R)$ changements de signe.
- (C) Si $\bar{d}(R) = m + n + 1$, il existe une constante $\beta > 0$ telle que :

$$\|\tau f - \tau g\| < \beta \|f - g\| \quad \forall g \in C_0[a, b]$$

où τf désigne l'élément optimal pour f .

Pour $n = 0$, on retrouve les résultats de l'approximation polynomiale avec des conditions de noeuds (PASZKOWSKI [1955], [1957] et dans le cas des noeuds extérieurs à l'intervalle: BERNSTEIN [1926]).

Pour $K = 0$, on retrouve l'approximation classique d'une fonction par une fraction rationnelle (voir par exemple ACHIESER [1956]).

Lorsqu'on prend $f(\xi_i) = \theta_i$ ($i = 1, \dots, K$), on obtient l'approximation par des fractions rationnelles égales à la fonction donnée en des points donnés (GILORMINI [1966], [1967], [1968]).

2 - Fractions rationnelles trigonométriques avec conditions de noeuds.

A tout polynôme trigonométrique $p(\theta)$, de degré m :

$$(3) \quad p(\theta) = \sum_{j=0}^m (a_j \cos j \theta + b_j \sin j \theta)$$

on associe, en posant $x = tg \frac{\theta}{2}$, un polynôme algébrique $p_1(x)$, de degré $2m$ au plus, défini par :

$$(4) \quad p(\theta) = \frac{p_1(x)}{(1+x^2)^m}$$

Nous dirons que deux polynômes trigonométriques $p(\theta)$ et $q(\theta)$ sont premiers entre eux si les polynômes algébriques $p_1(x)$ et $q_1(x)$ qu'on leur associe par la formule (4) sont premiers entre eux.

Problème : Prenons $I = [-\pi, +\pi]$ et soit $u(\theta)$ une fonction continue, de période 2π , strictement positive. Désignons par $\xi_1, \dots, \xi_K$ des noeuds contenus dans $]-\pi, +\pi]$. On cherche à approcher une fonction $f \in C[-\pi, +\pi]$ par la classe $\mathcal{W}_{m,n}$ des fonctions :

$$R(\theta) = u(\theta) \frac{p(\theta)}{q(\theta)} = u(\theta) \frac{\sum_0^m (a_j \cos j \theta + b_j \sin j \theta)}{\sum_0^n (c_j \cos j \theta + d_j \sin j \theta)}$$

telles que :

$$\begin{aligned} q(\theta) &> 0 \quad \text{sur} \quad [-\pi, +\pi] \\ p(\theta) \text{ et } q(\theta) &\text{ sont premiers entre eux} \\ R(\xi_i) &= \lambda_i \quad i = 1, \dots, K \end{aligned}$$

où $\lambda_1, \dots, \lambda_K$ sont des nombres réels donnés.

Existence d'un élément optimal : ici la classe $\mathcal{W}_{m,n}$ est l'ensemble des fonctions $R \in C[-\pi, +\pi]$ qui vérifie une identité de la forme :

$$(5) \quad \bar{q}(\theta) R(\theta) = u(\theta) \bar{p}(\theta)$$

où $\overline{p}(\theta)$ et $\overline{q}(\theta)$ sont deux polynômes trigonométriques à coefficients réels tels que :

$$\begin{cases} d^0(\overline{p}) \leq m & , & d^0(\overline{q}) \leq n \\ \overline{q}(\theta) \neq 0 \\ u(\xi_i) \overline{p}(\xi_i) - \lambda_i \overline{q}(\xi_i) = 0 & i = 1, \dots, K \end{cases}$$

Pour $R \in \overline{W}_{m,n}$ vérifiant (5), nous écrirons :

$$R = (\overline{p} \mid \overline{q})$$

On a tout d'abord (d'après le théorème IV-1), le résultat suivant :

Théorème 5 : toute fonction $f \in C[-\pi, +\pi]$ possède dans $\overline{W}_{m,n}$ au moins un élément optimal.

Remarque : soit $R = (\overline{p} \mid \overline{q})$ un élément de $\overline{W}_{m,n}$. Si on a :

$$\overline{q}(\xi_i) \neq 0 \quad i = 1, \dots, K$$

alors $R(x)$ peut être identifié avec un élément de $W_{m,n}$ comme le montre le lemme suivant :

Lemme : soient $p(\theta)$ et $q(\theta)$ deux polynômes trigonométriques tels que :

$$|p(\theta)| \leq |q(\theta)| \quad \forall \theta$$

Si $q(\theta_0) = 0$, il existe deux polynômes trigonométriques $p^*(\theta)$ et $q^*(\theta)$ tels que :

$$p^*(\theta) q(\theta) = p(\theta) q^*(\theta)$$

$$d^0(p^*) < d^0(p) \quad , \quad d^0(q^*) < d^0(q)$$

$$q^*(\theta_0) \neq 0$$

(CHENEY [1966], RIVLIN [1964]).

(a) Montrons tout d'abord que le polynôme algébrique $p_1(x)$ associé au polynôme trigonométrique $p(\theta)$ par la formule (4) a pour degré : $2d^0(p) - r$, où $r \geq 0$ désigne l'ordre de multiplicité de π comme zéro de $p(\theta)$.

En effet, pour $s \geq 0$, on peut écrire :

$$\frac{p_1(x)}{x^{2m-s}} = \frac{(1+x^2)^m}{x^{2m}} \frac{p(\theta)}{[tg \frac{\theta}{2}]^{-s}} = \frac{(1+x^2)^m}{x^{2m}} [\sin \frac{\theta}{2}]^s \frac{p(\theta)}{[\cos \frac{\theta}{2}]^s}$$

Lorsque $x \rightarrow +\infty$, on a : $\theta \rightarrow \pi$ et donc :

$$\lim_{x \rightarrow +\infty} \frac{p_1(x)}{x^{2m-s}} = \lim_{\theta \rightarrow \pi} \frac{p(\theta)}{[\cos \frac{\theta}{2}]^s}$$

D'où, pour $s \geq 1$, en appliquant la règle de l'Hopital :

$$\lim_{x \rightarrow +\infty} \frac{p_1(x)}{x^{2m-s}} = -\frac{2}{s} \lim_{\theta \rightarrow \pi} \frac{p'(\theta)}{[\cos \frac{\theta}{2}]^{s-1}} = \dots = \frac{(-2)^s}{s!} p^{(s)}(\pi)$$

ce qui prouve que : $d^0(p_1) = 2m - r$.

(b) Montrons ensuite que si $\theta_0 \neq \pi$ est zéro d'ordre r de $p(\theta)$, alors $x_0 = tg \frac{\theta_0}{2}$ est zéro d'ordre r du polynôme $p_1(x)$ associé à $p(\theta)$ par la formule (4)

On a en effet :

$$\begin{aligned} p_1(x) &= (1+x^2)^m p(2 \operatorname{Arc} tg x) \\ p_1'(x) &= (1+x^2)^{m-1} [2m x p + 2 p'] \\ &\dots\dots\dots \\ p^{(s)}(x) &= (1+x^2)^{m-s} [\alpha_0(x) p + \alpha_1(x) p' + \dots + \alpha_s(x) p^{(s)}] \end{aligned}$$

où $\alpha_0(x), \alpha_1(x), \dots, \alpha_s(x)$ sont des polynômes.

On en déduit :

$$p_1(x_0) = p_1'(x_0) = \dots = p_1^{(r-1)}(x_0) = 0$$

ce qui montre que x_0 est zéro d'ordre r au moins de $p_1(x)$.

Inversement, on a :

$$p(\theta) = \frac{p_1(x)}{(1+x^2)^m}, \quad x = tg \frac{\theta}{2}$$

$$p'(\theta) = \frac{1}{(1+x^2)^m} \left[\frac{1}{2} (1+x^2) p_1'(x) - 2mx p_1(x) \right]$$

.....

$$p^{(s)}(\theta) = \frac{1}{(1+x^2)^m} [\beta_0(x) p_1(x) + \beta_1(x) p_1'(x) + \dots + \beta_s(x) p_1^{(s)}(x)]$$

où $\beta_0(x), \beta_1(x), \dots, \beta_s(x)$ sont des polynômes.

Ceci montre que si $x_0 = tg \frac{\theta_0}{2}$ est zéro d'ordre r de $p_1(x)$, alors θ_0 est zéro d'ordre r au moins de $p(\theta)$.

C'est-à-dire que si $\theta_0 \neq \pi$ est zéro d'ordre r de $p(\theta)$, alors $x_0 = tg \frac{\theta_0}{2}$ est zéro d'ordre r de $p_1(x)$.

(c) Montrons maintenant que si $P_1(x)$ est un polynôme algébrique de degré $2N$ au plus, alors

$$P(\theta) = P_1(tg \frac{\theta}{2}) / (1 + tg^2 \frac{\theta}{2})^N$$

est un polynôme trigonométrique de degré N au plus.

En effet, $P_1(x)$ peut s'écrire :

$$\begin{aligned} P_1(x) &= \sum_{i=0}^N \alpha_{2i} x^{2i} + \sum_{i=0}^{N-1} \alpha_{2i+1} x^{2i+1} \\ &= \sum_{i=0}^N \beta_i (1+x^2)^{N-i} (1-x^2)^i + x \sum_{i=0}^{N-1} \beta_i' (1+x^2)^{N-1-i} (1-x^2)^i \end{aligned}$$

D'où, en posant $x = tg \frac{\theta}{2}$:

$$\begin{aligned} P(\theta) &= \sum_{i=0}^N \beta_i (\cos \theta)^i + \sum_{i=0}^{N-1} \beta_i'' \sin \theta (\cos \theta)^i \\ &= \sum_{j=0}^N \gamma_j \cos j \theta + \sum_{j=0}^{N-1} \gamma_j' \sin \theta \cos j \theta \\ &= \sum_{j=0}^N (\gamma_j \cos j \theta + \delta_j \sin j \theta) \end{aligned}$$

(d) Considérons maintenant deux polynômes trigonométriques $p(\theta)$ et $q(\theta)$ de degrés m et n tels que :

$$(6) \quad |p(\theta)| \leq |q(\theta)| \quad \forall \theta$$

$$(7) \quad q(\theta_0) = 0$$

La relation (6) montre que si θ_0 est zéro d'ordre $r \geq 1$ de $q(\theta)$, c'est un zéro d'ordre r au moins de $p(\theta)$.

Supposons $\theta_0 \neq \pi$ (si $\theta_0 = \pi$, on fait un changement de variable $\theta \rightarrow \theta + \alpha$) et posons :

$$x = tg \frac{\theta}{2}, \quad x_0 = tg \frac{\theta_0}{2}$$

$$p_1(x) = (1+x^2)^m p(\theta)$$

$$q_1(x) = (1+x^2)^n q(\theta)$$

D'après (a) et (b), on a :

$$d^0(p_1) \leq 2m, \quad d^0(q_1) \leq 2n$$

$$p_1(x) = (x-x_0)^r p_2(x)$$

$$q_1(x) = (x-x_0)^r q_2(x)$$

$$q_2(x_0) \neq 0$$

Si $r=1$, $q(\theta)$ étant périodique de période 2π , il existe un autre zéro θ_1 dans $[-\pi, +\pi]$ et on peut supposer $\theta_1 \neq \pi$. On a donc en posant $x_1 = tg \frac{\theta_1}{2}$:

$$p_1(x) = (x-x_0)^r (x-x_1)^{r_1} p_3(x)$$

$$q_1(x) = (x-x_0)^r (x-x_1)^{r_1} q_3(x)$$

$$q_3(x_0) \neq 0$$

$$r_1 \geq 0, \quad r+r_1 \geq 2$$

D'où l'on déduit :

$$d^0(p_3) \leq 2m-2, \quad d^0(q_3) \leq 2n-2$$

En posant :

$$p^*(\theta) = \frac{p_3(x)}{(1+x^2)^{m-1}}, \quad q^*(\theta) = \frac{q_3(x)}{(1+x^2)^{n-1}}$$

on a immédiatement :

$$p q^* - p^* q \equiv 0$$

$$q^*(\theta_0) \neq 0$$

$$\bar{d}^0(p^*) \leq m-1 < \bar{d}^0(p)$$

$$\bar{d}^0(q^*) \leq n-1 < \bar{d}^0(q)$$

ce qui achève la démonstration du lemme.

Propriété d'alternance : il faut calculer les nombres η^* et ν^* définis au chapitre précédent.

Soit $R \in W_{m,n}$:

$$R(\theta) = u(\theta) \frac{p(\theta)}{q(\theta)}$$

Posons : $d'(R) = 1 + 2 [m + n - \min \{m - \bar{d}^0(p), n - \bar{d}^0(q)\}]$

Tout élément $\Phi \in \mathcal{D}_R$ s'écrit :

$$\Phi(\theta) = \frac{u(\theta)}{q(\theta)} \Psi(\theta)$$

où
$$\Psi(\theta) = q(\theta) \sum_{j=0}^m (a_j' \cos j\theta + b_j' \sin j\theta) - p(\theta) \sum_{j=0}^n (c_j' \cos j\theta + d_j' \sin j\theta)$$

est un polynôme trigonométrique de degré au plus égal à :

$$m + n - \min \{m - \bar{d}^0(p), n - \bar{d}^0(q)\} = \frac{d'(R) - 1}{2}$$

tel que :

$$\Psi(\xi_i) = 0 \quad i = 1, \dots, K$$

Posons :

$$x = \operatorname{tg} \frac{\theta}{2}$$

$$\Psi_1(x) = \Psi(\theta) \times (1+x^2)^{\frac{d'(R)-1}{2}}$$

$\Psi_1(x)$ est un polynôme algébrique de degré : $2\bar{d}^0(\Psi) - r$, où r est l'ordre de multiplicité de π comme zéro de Ψ .

$\Psi(\theta)$ a donc au plus $d'(R) - 1$ zéros et donc :

$$\nu^*(\mathcal{D}_R) = d'(R)$$

On a d'autre part :

$$\dim(\mathcal{D}_R) = d'(R) - K = \nu^*(\mathcal{D}_R) - K$$

ce qui prouve que :

$$\eta^*(\mathcal{D}_R) = d'(R) - K$$

On en déduit alors des résultats analogues à ceux du paragraphe précédent :

Théorème 6 : soit $R \in W_{m,n}$. S'il existe un ensemble $Z = \{z_1, \dots, z_N\}$ de points de $[-\pi, +\pi]$, distincts des noeuds, tels que $y_R(z_i) \neq 0$ et si la suite $\tau(R, Z)$ présente au moins $d'(R)$ changements de signe, alors :

$$\rho \geq \min_{1 \leq i \leq N} |f(z_i) - R(z_i)|$$

Théorème 7 : pour que $R \in W_{m,n}$ soit optimal pour f , il faut et il suffit que l'une des conditions suivantes soit vérifiée :

$$(a) \quad \|f - R\| = \max_{1 \leq i \leq K} |f(\xi_i) - \lambda_i|$$

(b) Il existe un ensemble Z de (e)-points de $f - R$, distincts des noeuds, tel que la suite $\tau(R, Z)$ présente au moins $d'(R)$ changements de signe.

Théorème 8 : soit R optimal pour f dans $W_{m,n}$. Si on a :

$$\|f - R\| > \max_{1 \leq i \leq K} |f(\xi_i) - \lambda_i|$$

alors R est le seul élément optimal pour f dans $W_{m,n}$.

Enfin les théorèmes IV-13 (unicité forte) et IV-14 (continuité) subsistent en remplaçant l'hypothèse :

$$\nu^*(\mathcal{D}_{R^*}) = \eta^*(\mathcal{D}_{R^*}) + K = m + n + 1$$

par la condition :

$$d'(R) = 2m + 2n + 1 .$$

Remarque : comme au paragraphe précédent, considérons la classe $C_0[-\pi, +\pi]$ des fonctions $f \in C[-\pi, +\pi]$ telles que :

- a) $f(x)$ a au moins un élément optimal R^* dans $W_{m,n}$.
- b) cet élément vérifie :

$$\|f - R^*\| > \max_{1 \leq i \leq K} |f(\xi_i) - \lambda_i|$$

On a alors les résultats suivants :

- (A) Toute fonction $f \in C_0[-\pi, +\pi]$ a dans $W_{m,n}$ un élément optimal unique.
- (B) Cet élément est caractérisé par l'existence d'un ensemble Z de (e) -points tels que la suite $r(R, Z)$ présente au moins $d'(R)$ changements de signe.
- (C) Si $d'(R) = 2m + 2n + 1$, il existe $\beta > 0$ tel que :

$$\|\tau f - \tau g\| < \beta \|f - g\| \quad \forall g \in C_0[-\pi, +\pi]$$

3 - Fractions rationnelles ordinaires avec conditions d'inégalité.

Problème : soit $I = [a, b]$ un intervalle compact de la droite réelle. On cherche à approcher une fonction $f \in C[a, b]$ par la classe $W_{m,n}$ des fonctions

$$R(x) = u(x) \frac{p(x)}{q(x)}$$

où $u(x)$ est une fonction réelle donnée dans $C[a, b]$, strictement positive dans I et où $p(x), q(x)$ sont deux polynômes algébriques à coefficients réels tels que :

$$\begin{aligned} d^0(p) &\leq m, & d^0(q) &\leq n \\ q(x) &> 0 & \text{sur } [a, b] \\ p(x) &\text{ et } q(x) & \text{sont premiers entre eux} \\ V(x) &\leq R(x) \leq W(x) & \forall x \in [a, b] \end{aligned}$$

où $V(x)$, $W(x)$ sont deux fonctions données dans $C[a, b]$. On suppose-
ra qu'il n'existe aucun intervalle contenu dans $[a, b]$ dans lequel :

$$W(x) - V(x) \equiv 0$$

Remarquons que les contraintes se mettent également sous la forme :

$$V(x) \leq u(x) \leq W(x) \quad \forall x \in [a, b]$$

Existence d'un élément optimal : nous allons démontrer le résultat
suivant :

Théorème 9 : toute fonction $f \in C[a, b]$ possède dans $\bar{W}_{m, n}$ au moins
un élément optimal.

D'après le théorème II-1, nous savons que f possède au moins un
élément optimal dans la classe $\bar{W}_{m, n}$, ensemble des fonctions $R \in C[a, b]$
qui vérifient une identité de la forme :

$$(8) \quad q_1(x) R(x) \equiv u(x) p_1(x)$$

où $p_1(x)$, $q_1(x)$ sont deux polynômes algébriques à coefficients réels
tels que :

$$\begin{aligned} d^0(p_1) &\leq m, & d^0(q_1) &\leq n \\ q_1(x) &\equiv 0 \end{aligned}$$

$$(9) \quad V(x) \leq q_1(x) \leq W(x) \quad \forall x \in [a, b]$$

Nous noterons un élément $R \in \bar{W}_{m, n}$ sous la forme :

$$R = (p_1 \mid q_1)$$

Soit donc $R = (p_1 \mid q_1)$ optimal pour f dans $\bar{W}_{m, n}$. Nous allons montrer
que R s'identifie à un élément de $W_{m, n}$.

Remarquons tout d'abord que la relation (9) entraîne :

$$V(x) \leq q_1(x) \leq W(x) \quad \forall x \in [a, b]$$

ce qui, compte tenu de l'inégalité :

$$V(x) \leq W(x) \quad \forall x \in [a, b]$$

prouve que :

$$(10) \quad q_1(x) \geq 0 \quad \forall x \in [a, b]$$

D'autre part, la continuité de $R(x)$ montre que si x_0 est zéro d'ordre r de $q_1(x)$, alors x_0 est zéro d'ordre r au moins de $p_1(x)$. En outre, si $x_0 \in]a, b[$, la relation (10) montre que r est pair ($r = 2r'$).

Soit donc x_0 un zéro d'ordre r de $q_1(x)$. D'après ce qui précède, nous pouvons poser :

$$\begin{aligned} \text{si } x_0 \leq a : \\ q_1(x) &= (x - x_0)^r q_2(x) \\ p_1(x) &= (x - x_0)^r p_2(x) \end{aligned}$$

$$\begin{aligned} \text{si } x_0 \geq b : \\ q_1(x) &= (x_0 - x)^r q_2(x) \\ p_1(x) &= (x_0 - x)^r p_2(x) \end{aligned}$$

$$\begin{aligned} \text{si } a < x_0 < b : \\ q_1(x) &= (x - x_0)^{2r'} q_2(x) \\ p_1(x) &= (x - x_0)^{2r'} p_2(x) \end{aligned}$$

avec dans tous les cas : $q_2(x_0) \neq 0$

En divisant la relation (9) par $(x - x_0)^r$, on obtient alors :

$$V(x) q_2(x) \leq u(x) p_2(x) \leq W(x) q_2(x) \quad \forall x \in [a, b] - \{x_0\}$$

et lorsque $x_0 \in [a, b]$, cette relation est également vérifiée (par continuité) pour $x = x_0$.

Ce qui montre que l'on a :

$$R = (p_1 \mid q_1) = (p_2 \mid q_2)$$

avec : $q_2(x_0) \neq 0$.

Il est donc possible d'éliminer tous les zéros de $q_1(x)$ (et de la même manière d'éventuels facteurs du second degré diviseurs communs de p_1 et q_1). Ce qui donne :

$$R = (p_0 \mid q_0)$$

où p_0 et q_0 sont premiers entre eux et où $q_0(x)$ est strictement positif dans $[a, b]$. C'est-à-dire que $R \in W_{m,n}$, ce qui achève la démonstration du théorème.

Propriétés d'alternance : soit $R \in W_{m,n}$ optimal pour f . D'après les résultats du chapitre précédent (paragraphe 9), il existe $(s+s')$ points t_i de $[a,b]$ ($s \geq 0$, $s' \geq 0$) tels que :

$$R(t_i) = V(t_i) \quad i = 1, \dots, s$$

$$R(t_i) = W(t_i) \quad i = s+1, \dots, s+s'$$

Désignons encore par $F(R; t_1, \dots, t_{s+s'})$ la classe des fonctions :

$$\Phi(x) = u(x) p(x) - R(x) q(x)$$

telles que :

$$u(t_i) p(t_i) - V(t_i) q(t_i) = 0 \quad i = 1, \dots, s$$

$$u(t_i) p(t_i) - W(t_i) q(t_i) = 0 \quad i = s+1, \dots, s+s'$$

ou encore :

$$\Phi(t_i) = 0 \quad i = 1, \dots, s+s'$$

D'après ce que nous avons vu au paragraphe 1, on a :

$$\nu^*[F(R; t_1, \dots, t_{s+s'})] = d(R)$$

$$\eta^*[F(R; t_1, \dots, t_{s+s'})] = d(R) - (s+s')$$

ce qui prouve que $F(R; t_1, \dots, t_{s+s'})$ vérifie la condition de Haar sur $[a,b] - \{t_1, \dots, t_{s+s'}\}$.

D'après le chapitre précédent, on a donc :

Théorème 10 : soit $R(x)$ optimal pour f dans $W_{m,n}$

(a) Il existe $(s+s')$ points t_i de $[a,b]$, avec $s \geq 0$, $s' \geq 0$, tels que :

$$R(t_i) = V(t_i) \quad i = 1, \dots, s$$

$$R(t_i) = W(t_i) \quad i = s+1, \dots, s+s'$$

(b) Si aucun des points t_i n'est (e)-point de $f-R$, alors il existe $[d(R) + 1 - (s+s')]$ (e)-points de $f-R$, distincts des points t_i , tels qu'entre deux (e)-points consécutifs de même

genre il y a un nombre impair de points t_i et entre deux (e)-points consécutifs de genres différents un nombre pair (ou nul) de points t_i .

(a) Si aucun des points t_i n'est (e)-point de $f-R$, alors R est le seul élément optimal pour f dans $W_{m,n}$.

Remarque : Désignons par $R_{m,n}$ la classe des fonctions rationnelles :

$$R(x) = \frac{p(x)}{q(x)} = \frac{a_0 + a_1 x + \dots + a_m x^m}{b_0 + b_1 x + \dots + b_n x^n}$$

telles que $q(x) > 0$

et posons :

$$r = \inf_{R_{m,n}} \|f - R\|$$

Si on prend les fonctions $V(x)$ et $W(x)$ telles que :

$$W(x) > f(x) + r$$

$$V(x) < f(x) - r$$

on a alors : $s = s^* = 0$ et on retrouve à l'aide du théorème 10, le résultat classique :

il existe dans $R_{m,n}$ un élément optimal unique R pour f et le nombre de (e)-points alternés de $f-R$ est au moins égal à $[d(R) + 1]$.

4 - Fractions rationnelles trigonométriques avec conditions d'inégalité.

Ici encore comme au paragraphe 2, à tout polynôme trigonométrique $p(\theta)$ de degré m :

$$p(\theta) = \sum_{j=0}^m (a_j \cos j \theta + b_j \sin j \theta)$$

nous associons, en posant $x = tg \frac{\theta}{2}$ un polynôme algébrique $p_1(x)$ de

degré $2m$ au plus, défini par :

$$(11) \quad p(\theta) = \frac{p_1(x)}{(1+x^2)^m}$$

et nous dirons que deux polynômes trigonométriques $p(\theta)$ et $q(\theta)$ sont premiers entre eux si les polynômes algébriques $p_1(x)$ et $q_1(x)$ qu'on leur associe par la formule (11) sont premiers entre eux.

Problème : prenons $I = [-\pi, +\pi]$ et désignons par $u(\theta)$ une fonction continue, de période 2π , strictement positive. On cherche à approcher une fonction $f \in C[-\pi, +\pi]$ par la classe $W_{m,n}$ des fonctions :

$$R(\theta) = u(\theta) \frac{p(\theta)}{q(\theta)} = u(\theta) \frac{\sum_0^m (a_j \cos j\theta + b_j \sin j\theta)}{\sum_0^n (c_j \cos j\theta + d_j \sin j\theta)}$$

telles que :

$$q(\theta) > 0 \quad \text{sur} \quad [-\pi, +\pi]$$

$$p(\theta) \text{ et } q(\theta) \text{ sont premiers entre eux}$$

$$V(\theta) \leq R(\theta) \leq W(\theta) \quad \forall \theta \in [-\pi, +\pi]$$

où $V(\theta)$, $W(\theta)$ sont deux fonctions données dans $C[-\pi, +\pi]$.

On supposera (comme au paragraphe précédent) qu'il n'existe aucun intervalle contenu dans $[-\pi, +\pi]$ sur lequel on ait :

$$V(\theta) - W(\theta) \equiv 0$$

Existence d'un élément optimal : nous allons démontrer le résultat suivant :

Théorème 11 : toute fonction $f \in C[-\pi, +\pi]$ possède dans $W_{m,n}$ au moins un élément optimal.

Nous savons déjà (théorème II-1) que f possède au moins un élément optimal dans la classe $\bar{W}_{m,n}$, ensemble des fonctions $R \in C[-\pi, +\pi]$

qui vérifient une identité de la forme :

$$(12) \quad q(\theta) R(\theta) \equiv u(\theta) p(\theta)$$

où $p(\theta)$ et $q(\theta)$ sont deux polynômes trigonométriques à coefficients réels tels que :

$$d^0(p) \leq m, \quad d^0(q) \leq n$$

$$q(\theta) \not\equiv 0$$

$$(13) \quad V(\theta) q(\theta) \leq u(\theta) p(\theta) \leq W(\theta) q(\theta) \quad \forall \theta \in [-\pi, +\pi]$$

Nous noterons un élément $R \in \widetilde{W}_{m,n}$ sous la forme :

$$R = (p | q)$$

Soit donc $R = (p | q)$ optimal pour f dans $\widetilde{W}_{m,n}$.

Remarquons que la relation (13) montre que l'on a :

$$(14) \quad q(\theta) \geq 0 \quad \forall \theta \in [-\pi, +\pi]$$

En outre, la continuité de $R(\theta)$ prouve que si θ_0 est zéro d'ordre r de $q(\theta)$, c'est un zéro d'ordre r au moins de $p(\theta)$.

Supposons d'abord que $q(\theta)$ s'annule pour $\theta_0 \in]-\pi, +\pi[$. D'après (14), θ_0 est un zéro d'ordre $2r$ (r entier). En posant $x = \operatorname{tg} \frac{\theta}{2}$, nous associons aux polynômes trigonométriques $p(\theta)$ et $q(\theta)$ deux polynômes algébriques $p_1(x)$ et $q_1(x)$ tels que :

$$(15) \quad p(\theta) = \frac{p_1(x)}{(1+x^2)^m}, \quad q(\theta) = \frac{q_1(x)}{(1+x^2)^n}$$

D'après ce que nous avons vu au paragraphe 2, $x_0 = \operatorname{tg} \frac{\theta_0}{2}$ est zéro d'ordre $2r$ pour $q_1(x)$ et d'ordre au moins $2r$ pour $p_1(x)$, ce qui permet d'écrire :

$$p(\theta) = \frac{(x-x_0)^{2r} p_2(x)}{(1+x^2)^m}$$

$$q(\theta) = \frac{(x-x_0)^{2r} q_2(x)}{(1+x^2)^n}$$

avec :

$$q_2(x_0) \neq 0$$

La relation (13) devient alors :

$$(16) \quad \frac{V(\theta) q_2(x) (x-x_0)^{2r}}{(1+x^2)^n} \leq \frac{u(\theta) p_2(x) (x-x_0)^{2r}}{(1+x^2)^m} \leq \frac{W(\theta) q_2(x) (x-x_0)^{2r}}{(1+x^2)^n}$$

Posons alors :

$$\bar{p}(\theta) = \frac{p_2(x)}{(1+x^2)^m}, \quad \bar{q}(\theta) = \frac{q_2(x)}{(1+x^2)^n}$$

$\bar{p}(\theta)$ et $\bar{q}(\theta)$ sont deux polynômes trigonométriques tels que :

$$\bar{q}(\theta) R(\theta) = u(\theta) \bar{p}(\theta)$$

$$d^0(\bar{p}) \leq m, \quad d^0(\bar{q}) \leq n$$

$$\bar{q}(\theta_0) \neq 0$$

et, en divisant les deux membres de (16) par $(x-x_0)^{2r}$:

$$\bar{q}(\theta) V(\theta) \leq \bar{p}(\theta) u(\theta) \leq \bar{q}(\theta) W(\theta) \quad \forall \theta \in [-\pi, +\pi]$$

c'est-à-dire que l'on a :

$$R = (\bar{p} | \bar{q}) \quad \text{avec} \quad \bar{q}(\theta_0) \neq 0$$

On peut donc de cette manière éliminer tout zéro de $\bar{q}(\theta)$ appartenant à $]-\pi, +\pi[$.

Supposons maintenant que l'élément $R = (p | q)$ de $\tilde{W}_{m,n}$ soit tel que π soit zéro de $q(\theta)$. D'après la relation (14) et la périodicité, on voit que c'est nécessairement un zéro d'ordre pair $2r$. On a donc (d'après le paragraphe 2) :

$$p(\theta) = \frac{p_1(x)}{(1+x^2)^m}, \quad q(\theta) = \frac{q_1(x)}{(1+x^2)^n}$$

avec :

$$d^0(q_1) = 2n - 2r, \quad d^0(p_1) \leq 2m - 2r$$

On pose alors :

$$\bar{p}(\theta) = \frac{p_1(x)}{(1+x^2)^{m-r}}, \quad \bar{q}(\theta) = \frac{q_1(x)}{(1+x^2)^{n-r}}$$

et on vérifie comme ci-dessus que l'on a :

$$R = (\overline{p} \mid \overline{q}) \quad \text{avec} \quad \overline{q}(\pi) \neq 0$$

Nous venons donc de montrer que :

$$R = (\overline{p} \mid \overline{q})$$

avec : $q(\theta) > 0 \quad \text{sur} \quad [-\pi, +\pi]$

c'est-à-dire que $R \in W_{m,n}$ ce qui achève la démonstration du théorème 11.

Propriétés d'alternance : soit $R \in W_{m,n}$ optimal pour f . On a vu (chapitre IV, paragraphe 9) qu'il existe alors $(s+s')$ points θ_i ($s \geq 0, s' \geq 0$) dans $[-\pi, +\pi]$ tels que :

$$\begin{aligned} R(\theta_i) &= V(\theta_i) & i &= 1, \dots, s \\ R(\theta_i) &= W(\theta_i) & i &= s+1, \dots, s+s' \end{aligned}$$

Désignons encore par $F(R; \theta_1, \dots, \theta_{s+s'})$ la classe des fonctions

$$\Phi(\theta) = u(\theta) p(\theta) - R(\theta) q(\theta)$$

où :

$$p(\theta) = \sum_0^m (a_j \cos j \theta + b_j \sin j \theta)$$

$$q(\theta) = \sum_0^n (c_j \cos j \theta + d_j \sin j \theta)$$

$$\Phi(\theta_i) = 0 \quad i = 1, \dots, s+s'$$

D'après le paragraphe 2, on a :

$$\nu^*[F(R; \theta_1, \dots, \theta_{s+s'})] = d^+(R)$$

$$\eta^*[F(R; \theta_1, \dots, \theta_{s+s'})] = d^+(R) - (s+s')$$

avec : $d^+(R) = 1 + 2 [m+n - \min(m - d^0(p), n - d^0(q))]$

Ceci prouve que $F(R; \theta_1, \dots, \theta_{s+s'})$ vérifie la condition de Haar sur $[-\pi, +\pi] - \{\theta_1, \dots, \theta_{s+s'}\}$. On a donc d'après le chapitre précédent le résultat suivant :

Théorème 12: soit $R(\theta)$ optimal pour f dans $W_{m,n}$

- (a) Il existe $(s+s')$ points θ_i de $[-\pi, +\pi]$, avec $s \geq 0$, $s' \geq 0$ tels que:

$$\begin{aligned} R(\theta_i) &= V(\theta_i) & i &= 1, \dots, s \\ R(\theta_i) &= W(\theta_i) & i &= s+1, \dots, s+s' \end{aligned}$$

- (b) Si aucun des points θ_i n'est (e)-point de $f-R$, alors il existe $[\bar{d}'(R) + 1 - (s+s')]$ (e)-points de $f-R$, distincts des points θ_i , tels qu'entre deux (e)-points consécutifs de même genre il y a un nombre impair de points θ_i et entre deux (e)-points consécutifs de genres différents un nombre pair (ou nul) de points θ_i .

- (c) Si aucun des points θ_i n'est (e)-point de $f-R$, alors R est le seul élément optimal pour f dans $W_{m,n}$.

Remarque: Désignons par $T_{m,n}$ la classe des fractions trigonométriques:

$$R(\theta) = \frac{\sum_0^m (a_j \cos j \theta + b_j \sin j \theta)}{\sum_0^n (c_j \cos j \theta + d_j \sin j \theta)}$$

telles que:

$$\sum_0^n (c_j \cos j \theta + d_j \sin j \theta) > 0 \quad \text{sur} \quad [-\pi, +\pi]$$

et posons:

$$r = \inf_{T_{m,n}} \|f - R\|$$

Si on prend les fonctions $V(\theta)$ et $W(\theta)$ telles que:

$$W(\theta) > f(\theta) + r$$

$$V(\theta) < f(\theta) - r$$

on a alors $s=s'=0$ et on retrouve à l'aide du théorème 12, le résultat démontré par LOEB [1966]:

il existe dans $T_{m,n}$ un élément optimal unique R pour f et le nombre de (e)-points alternés de $f-R$ est au moins égal à $[\bar{d}'(R) + 1]$.

BIBLIOGRAPHIE

- ACHIESER [1956] - Theory of approximation, New York.
- BERNSTEIN [1926] - Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle, Paris.
- BOEHM [1965] - Existence of best rational Tchebycheff approximations, Pacific Journal of Mathematics, 15, 1, p. 19-28.
- BROSOWSKI [1965] - über Tchebysheffsche Approximationen mit linearen Nebenbedingungen, Math. Zeitschrift, 88, p. 105-128.
- BROSOWSKI [1965a] - über Tchebysheffsche Approximationen mit verallgemeinerten rationalen Funktionen, Math. Zeitschrift, 90, p. 140-151.
- BROSOWSKI [1965b] - über die Eindeutigkeit der rationalen Tchebysheff Approximationen, Numerische Mathematik, 7, p. 176-186.
- BUROV [1961] - The approximation of functions by polynomials satisfying non-linear relations, Doklady Akad. Nauk SSSR, 138, p. 515-517
= Soviet Mathematics, 2, p. 640-642.
- CHENEY [1964] - Approximation by generalized rational functions, Symposium on approximation of functions, Warren, Michigan (Garabédian, éditeur).
- CHENEY [1966] - Introduction to Approximation Theory, New-York.
- CHENEY et LOEB [1964] - Generalized rational approximation, Journal S.I.A.M., Numerical Analysis, 1, p. 11-25.
- CHENEY et LOEB [1966] - On the continuity of rational approximation operator, Archive for Rational Mechanics and Analysis, 21, p. 391-401.
- COLLATZ [1960] - Tchebysheffsche Annäherung mit rationalen Funktionen, Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg, 24, p. 70-78.

- COLLATZ [1964] - Funktionanalysis und Numerische Mathematik, Berlin.
- EGGLESTON [1958] - Convexity, Cambridge University Press.
- GILORMINI [1966] - Approximation rationnelle avec des noeuds, Comptes rendus, Académie des Sciences, Paris, 263, série A, p. 286-287.
- GILORMINI [1967] - Continuité de l'approximation rationnelle de Tchebycheff avec des noeuds, Comptes rendus, Académie des Sciences, Paris, 264, série A, p. 359-360.
- GILORMINI [1967 a] - Approximation par fractions rationnelles généralisées dont les coefficients vérifient des relations linéaires, Comptes rendus, Académie des Sciences, Paris, 264, série A, p. 795-798.
- GILORMINI [1967 b] - Sur l'approximation par fractions rationnelles généralisées avec contraintes sur les coefficients, Comptes rendus, Académie des Sciences, Paris, 265, série A, p. 253-256.
- GILORMINI [1968] - Sur un contre-exemple de M. Henry L. Loeb, Comptes rendus, Académie des Sciences, Paris, 266, série A, p. 239-241.
- GOLSHTEIN [1961] - Sur une généralisation des problèmes de Tchebysheff, Uspehi Mat. Nauk, 16, 4(100), p. 208.
- GOLSHTEIN [1961 a] - An infinite dimensional analogue of the problem of linear programming and its applications to some problems in the theory of approximation, Doklady Akad. Nauk SSSR, 140, 1, p. 23-26 = Soviet Mathematics, 2, p. 1125-1129.
- GOLSHTEIN [1962] - On a general formulation of the problem of best approximation, Doklady Akad. Nauk SSSR, 144, 1, p. 21-22 = Soviet Mathematics, 3, p. 637-639.
- KARLIN et STUDDEN [1966] - Tchebysheff systems with applications in analysis and statistics, New-York.
- LAURENT [1967] - Approximation uniforme de fonctions continues sur un compact avec contraintes de type inégalité, Revue Française d'Informatique et de Recherche Opérationnelle, n°5, p. 81-95.

- LOEB [1966] - Approximation by generalized rationals, Journal S.I.A.M., Numerical Analysis, 3, 1, p. 34-55.
- LOEB [1968] - Un contre exemple à un résultat de M. Claude Gilormini, Comptes rendus, Académie des Sciences, Paris, 266, série A, p. 237.
- MEINARDUS [1964] - Approximation von Funktionen und ihre numerische Behandlung, Berlin.
- PASZKOWSKI [1955] - Sur l'approximation uniforme avec des noeuds, Annales Polonici Mathematici, 2, p. 118-135.
- PASZKOWSKI [1957] - On approximation with nodes, Rozprawy Matematyczne, XIV, Varsovie.
- RICE [1963] - Approximation with convex constraints, Journal S.I.A.M., 11, 1, p. 15-32.
- RICE [1964] - The approximation of functions: linear theory, New-York.
- RIVLIN [1964] - A property of the ratio of trigonometric polynomials, Journal S.I.A.M., Numerical Analysis, 1, p. 131-132.
- SINGER [1956] - Caractérisation des éléments de meilleure approximation dans un espace de Banach quelconque, Acta Scientiarum Mathematicarum, Szeged, 17, p. 181-189.
- SINGER [1956 a] - Sur l'extension des fonctionnelles linéaires, Revue roumaine de Mathématiques Pures et appliquées, 1, 2, p. 99-106.
- SINGER [1957] - Asupra unicitatii elementului de cea mai buna aproximatie in spatii Banach barecare, Studii si cercet. mat., Bucarest, 8, p. 234-244.
- SINGER [1959] - La meilleure approximation interpolative dans les espaces de Banach, Revue roumaine de Mathématiques Pures et appliquées, 4, p. 95-113.

TERRENOIRE [1967] - Application de la programmation linéaire et convexe à l'approximation au sens de Tchebycheff avec contraintes, thèse de 3^{ème} cycle, Grenoble.

WERNER [1964] - On the rational Tchebysheff operator, Math. Zeitschrift, 86, p. 317-326.

ERRATA

Dans les résultats suivants :

Corollaire 3 du théorème II-5 (p. 59),

Théorème III-7 (p. 75),

Théorème IV-6 (p. 100),

il y a lieu de rajouter la condition suivante :

$$L(\Phi) = \sum_{i=1}^r \lambda_i \Phi(x_i) \quad \text{avec} \quad \sum |\lambda_i| = 1 .$$

DEUXIÈME THÈSE

Propositions données par la Faculté

" Théorèmes asymptotiques dans les chaînes de Markov."

Vu et approuvé

Paris, le ... 8 Mai 19 68.

Le Doyen de la Faculté des Sciences
de Paris

Marc. ZAMANSKY

Vu
et permis d'imprimer

Le Recteur de l'Académie de Paris

Jean ROCHE

